# Controlling fluxes for microbial metabolic engineering

# Controlling fluxes for microbial metabolic engineering

a dissertation presented

by

Gairik Sachdeva

to

The School of Engineering and Applied Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Engineering Sciences

Harvard University

Cambridge, Massachusetts

July 2014

Thesis advisor: Professor Pamela A. Silver Gairik Sachdeva

# Controlling fluxes for microbial metabolic engineering

## Abstract

This thesis presents novel synthetic biology tools and design principles usable for microbial metabolic engineering. Controlling metabolic fluxes is essential for biological manufacturing of fuels, materials, and high value chemicals. Insulating the flow of metabolites is a successful natural strategy for metabolic flux regulation. Recently, approaches using scaffolds, both *in vitro* and *in vivo*, to spatially co-localize enzymes have reported significant gains in product yields. RNA is suitable for use as an *in vivo* scaffold due to its diversity of structural motifs, predictability of design, and ease of expression.

We used scaffolds made from RNA strands containing two sets of motifs. Polymerization domains bring individually expressed strands together as tiles to form a macromolecular assembly. Aptamers included in the sequence recruit RNA binding domains to these *in vivo* scaffolds. We assembled a library of eight aptamers and corresponding RNA binding domains fused to partial fragments of fluorescent proteins in *E. coli*. New scaffold designs could co-localize split green fluorescent protein fragments to produce activity as measured by cell-based fluorescence. The scaffolds consisted of either single bivalent RNAs or RNAs designed to polymerize in 1 or 2 dimensions.

The new scaffolds were used to increase metabolic output from a two-enzyme pentadecane production pathway that contains a fatty aldehyde intermediate, as well as

three and four enzymes in the succinate production pathway. Pentadecane synthesis depended on the geometry of enzymes on the scaffold, as determined through systematic reorientation of the acyl-ACP reductase fusion by rotation via addition of base pairs to its cognate RNA aptamer. Together, these data suggest that intra-cellular scaffolding of enzymatic reactions may enhance the direct channeling of a variety of chemical substrates.

In the cyanobacterium *S. elongatus* we showed that the flow of electrons can be controlled and insulated by co-expression of a heterologous hydrogenase with different interacting ferrodoxin proteins. We used the system to demonstrate photo biological hydrogen production and hydrogen dependent cell growth. Finally, we expressed RNA scaffolds in *S. elongatus* using a T7 RNA polymerase system, opening doors for spatial control of metabolism in the cyanobacterium.

# Contents

# Listing of figures

For my loving family.

# Acknowledgments

I OWE MY GRATITUDE to several people who have helped and supported me during the journey through this Ph.D. program. I am thankful to Pam Silver and Jeff Way for guiding me as a part of their very innovative research group at Harvard. Their drive for being at the very edge of synthetic biology and encouragement of extraordinary research proposals was inspiring.

Pam also recruited some of the best scientists, whom I had the fortune of working with. Thank you, Danny Ducat and Buz Barstow for your mentorship and inspiration. Danny was the most patient and careful teacher, and the perfect mentor for an engineer being introduced to biology. Buz, with his encouragement and example, taught me to be an optimist, which I believe is one of the most helpful virtues a scientist can possess.

I am grateful to have collaborated with excellent scientists in Danny, Abhishek and Cameron, who have directly helped me put together the body of research presented here. Camille and Faisal, with their pioneering work on *in vivo* RNA scaffolds in the lab and engaging discussions, motivated a bulk of the work here. I also throughly en-

joyed learning a great deal about teaching and research when mentoring Simon and David as parts of their training towards becoming successful scientists.

The Silver lab was an enormous resource of knowledge, fun, cultural and outdoors experiences - all of which made me feel at home and made graduate school extremely enjoyable. Thank you Qingqing, Tyler, Mara, Patrick, Joe, Steph, Jordan, and the many dynamic constituents of the Silver lab.

I was fortunate to have wonderful friends and classmates make my time at Harvard truly memorable. I am glad to have shared some of the best times with Mike, Jean, Onur, Willy, Shradha, Janaki, AJ, Zsofia, Camille and many more. Thank you Krista, for unending support and the many delightful escapes.

Lastly, my family owes the bulk of my thanks, which I can't express enough of. My grandparents, uncles, aunts, cousins have always showered so much love. My parents, Bhawna and Sushil, and my sister Mehak have been the greatest sources of support, encouragement, inspiration, and unending love - all of which have been instrumental in helping cross every step on the way to this dissertation, and in life.

*Indeed, our lack of understanding of metabolic regulation has been revealed by poor results from attempts to increase the rates of selected metabolic pathways... Our ability to interfere with an organism's genetics has far outstripped our ability to predict the effects on its metabolism*

David Fell, *Understanding the Control of Metabolism*[1]

# 1

# Synthetic biology approaches for control of metabolism

Synthetic biology is rapidly advancing towards engineering at the nanoscale. The ability to engineer cells with sub-cellular spatial precision is a very powerful and essential tool for synthetic biologists, who look at cells as programmable

devices[2]. Such an approach aims to engineer biological systems to diagnose and treat diseases[3], perform molecular computations[4], and manufacture materials as parts of bio-refineries of the future[5]. The latter goal is especially urgent since biological replacements for petroleum based products can go long way in tackling the climate change problem[6].

Efforts in metabolic engineering have benefited from the synthetic biology approach of treating genetic and cellular components as modules that can be engineered to function together in novel ways[7]. However, understanding the complexity of metabolic regulation and going beyond that to engineer metabolism for the production of chosen compounds has been a long standing challenge[1]. Although there have been significant advances in both computational[8] and empirical[9] methods to study and design metabolism, significant challenges remain with respect to the catalytic rates of heterologously expressed enzymes within a given cell's metabolism[10]. One traditional approach has been to over produce the enzymes themselves; however, that is coupled with inefficiencies due to slower growth rates[10]. Other means to increase specific metabolic reaction yield include protein engineering[11], and sub-cellular co-localization of enzymes[12].

## 1.1 Spatial control for metabolic engineering

Co-localization of proteins, DNA and RNA enhances metabolic output of enzymes[13,14], allows novel regulation of gene expression[15,16,17], and can increase the specificity of therapeutics[18,19]. This occurs primarily because co-localized macromolecules have high local concentrations, allowing their activities to be coordinated. Thus, better ability to organize proteins, RNAs, lipids etc. into synthetic supramolecular complexes should enable diverse and more complex function than can be achieved by solely engineering

individual parts.

Spatial organization of biochemical reactions is commonly observed in biological systems and allows for complex metabolic control. Natural approaches include compartmentalization, separation of reactions in different cells, and supramolecular organizations of enzymes[20] These strategies control the direction of metabolic pathway flux[20,21], prevent loss of intermediates to competing cellular reactions[22], and insulate the cell itself from intermediates that may be volatile[23]. Polyketide and fatty acid synthases are examples of large multi-enzyme complexes where intermediates are shuttled between enzymes to achieve end products with specific lengths and chemical groups[24]. Similarly in bacteria, protein microcompartments for propanediol and ethanolamine utilization restrict the diffusion of toxic intermediates such as propionaldehyde and acetaldehyde respectively[25].

Co-localization of enzymes has been successfully employed in synthetic systems as well[22,26]. Confinement of enzymatic reactions has been demonstrated both *in vitro*[23,27,28,29] and *in vivo*. In bacteria, both protein[14,30] and DNA[31] have been used as scaffolds for heterologous enzymatic pathways with several fold increases in end product titers reported. Recently, our lab showed that RNA scaffolds can be used to modularly organize proteins in *E. coli*[13] and increase the rate of electron transfer between enzymes. Such scaffolds, formed by tiles[13] or repeats[32] of small RNA strands, lead to higher densities of binding sites available for protein docking.

There has been debate about the mechanism by which scaffolds enable metabolic substrate channeling. The transfer of electrons between enzymes relies on physical contact and thus is limited by protein diffusion rates and competition, which are effectively addressed by scaffolding[13]. However, the role of enzyme co-localization in pathways involving diffusible intermediates is much less well understood[33,12]. In a re-

cent study[28], enzymes localized in close proximity, less than 20 nanometers apart, on *in vitro* assembled DNA scaffolds exhibited enhanced rates of metabolite exchange. The transfer rates dropped precipitously with any further increase in inter-enzyme distance. Since such effects are not explicable by 3D diffusion models[33], a mechanism of metabolite substrate channeling by restricted diffusion on hydration layers across crowded protein surfaces has been proposed[28]. RNA scaffolds, with their predictable geometry, can be used to create a range of metabolic channeling platforms and test the relative effects from these two different mechanisms.

## 1.2 RNA for *in vivo* spatial engineering

Nucleic acids, which have important structural and catalytic roles in nature, are also enabling synthetic biology efforts for *in vivo* spatial engineering. Growing knowledge of the rules determining RNA structure and catalysis has led to the development of tools that allow one to both predict function of RNA from its sequence and rationally design RNA to assume desired motifs and interactions. Other advances have allowed RNA to be synthesized, studied, and manipulated both *in vitro* and *in vivo*. Here, we[*] present an account of these developments and how synthetic RNA scaffolds are advancing basic research, metabolic engineering and even therapeutics through *in vivo* spatial engineering. We offer a strong prognosis for the role of synthetic RNA devices as bioengineering tools with emerging synergies between RNA, DNA, protein nanotechnology and high-throughput sequencing.

Natural non-coding RNAs already play structural and catalytic roles in cells. A breadth of studies has established design principles that can be used to predictably

[*]This section forms part of a review, co-written with Cameron Myhrvold and Peng Yin, and submitted to Wiley Biotechnology Series

shape RNA secondary structures[34,35,36,37]. Structural malleability of RNA, the ease of expressing synthetic RNA constructs in cells, their stability and advances in methods for assaying and imaging assembled structures are some of the many reasons why RNA is a useful scaffolding material. Synthetic biology efforts have demonstrated that carefully designed RNA can be used for sub-cellular targeting of probes, enzymes, and therapeutic agents.

### 1.2.1 Structural Roles of Natural RNA

RNAs perform numerous biological functions as canonical gene expression agents, catalysts, gene regulation switches, and structural scaffolds. These structural and catalytic roles of RNA are due in large part to the tremendous diversity of secondary and tertiary structures assumed by natural RNA, and the fact that ribose sugars are more reactive than deoxyribose. RNA secondary structures can include intricate motifs like double helices, hairpin loops, bulges, pseudo-knots, and right-angled turns[38,39]. Aside from the Watson-Crick base pairing, RNA has the capacity to form Hoogsteen base pairs as well as wobble base pairs. Such interactions allow motifs to be connected in higher order tertiary interactions, pre-dominantly through the non Watson-Crick base pairs[40,41].

#### RNA as a natural catalyst

Catalytic roles of RNA during translation, like the tRNA, disrupted a simple view held by the central dogma that RNA exists merely to transfer genetic information from DNA to protein. Today we know that RNA has catalytic and regulatory roles in many other cellular processes as well. Regulatory RNA structures play a significant role in the

control of translation initiation of several bacterial genes and in bacterial immunity[42]. RNAs affect expression in *cis*, by forming secondary structures near translation start sites of the mRNA. The *cis* regions can bind to regulatory proteins or other RNAs that affect translation in *trans*[42]. Other similarly dynamic regulatory RNA regions can consist of aptamers, which are nucleic acids that selectively bind ligands[43]. Many metabolic genes are 'switched' on or off, triggered by the binding of small molecule metabolites to some of these regulatory RNAs known as riboswitches[44].

## RNA scaffolds in nature

There are also several instances of natural RNAs that are largely structural. Some natural RNAs are known to specifically bind the coat proteins of single-strand RNA phages. Such interactions help package the RNA into viral capsids. Some RNA phages that have well-characterized RNA binding proteins include PP7[45], MS2[46] and Qβ[47]. These coat proteins also act as repressors of the viral replicase translation by specifically binding RNA hairpins near the origin of replication. In the bacteriophage, ϕ29, a short (117-174 nt) sequence of packaging RNA (pRNA) helps to pack phage DNA into pre-formed capsids[48]. A DNA packaging motor is composed of a pentameric ring of pRNA, capsid proteins, ds-DNA and an ATPase[49]. Studies characterizing the specificity and stoichiometries of these interactions[45,50,51,52] have laid the foundation for RNA-tagging based applications.

RNA scaffolds are important in eukaryotic gene expression as well. Mammalian cells appear to extensively employ long noncoding RNAs. These lncRNAs are rich with secondary structure motifs[53,54] some of which bind and coordinate proteins on scaffolds that play important roles in epigenetic regulation[55,56] and telomere maintenance[57,58]. Thus, natural RNA diversity offers a template of diverse structure and

function for synthetic biologists. Next, we look at how natural observations have been translated into precise engineering of the structure and dynamics of RNA for *in vivo* applications.

### 1.2.2 Applications of designed RNA scaffolds

The design principles for RNA i.e. its structural and folding properties are well understood. Dynamic properties of RNA have been engineered to create synthetic riboregulators[59], ribozymes[60], and riboswitches[61,62]. Additionally, the self-assembly of RNA *in vitro* has also been well characterized and controlled to create precisely designed nano-structures like cubes[63], jigsaws[64], squares[65], and polyhedra[66]. RNA sequences consisting of secondary structures and functional units designed using these empirical results and computational tools[34,35,36,37] can be genetically expressed in cells. Such engineered RNAs have been used for tasks ranging from studying natural RNA processing in cells to metabolic engineering and therapeutic applications.

#### Tools for RNA research

While mRNA has long been known to function as a template for protein translation, the spatio-temporal aspects of the various steps involved in mRNA processing remain poorly understood. Investigation of the dynamics of mRNA as it goes through translation, splicing, nuclear export in eukaryotes, localization for translation, and finally degradation requires tools to track individual RNA molecules. Aptamers and their recruitment of fluorescent proteins on engineered mRNA scaffolds have enabled such studies. Some of the earliest attempts to tag RNA *in vivo* were carried out by expressing GFP fused with bacteriophage MS2 coat protein[67] or human RNA interacting

protein domain U1A[68] along with mRNA containing the corresponding binding sites in *Saccharomyces cerevisiae*. Such tags enabled tracking of single-cell mRNA localization by microscopy. Furthermore, by incorporating tandem repeats of MS2 binding sites on reporter mRNA[69], several GFP-MS2 fusions could be localized on a single transcript, enabling tracking of individual mRNA molecules in mammalian cells. This *in vivo* tracking method was extended to other systems[70], including bacteria[32,71].

More recently, several efforts have addressed the longstanding question of whether or not RNA is highly localized within bacterial cells[72,73]. A significant innovation over the above strategy came from the use of fluorescent protein complementation. In this approach, RNA aptamers are used to bring together two different protein fusion units, each with a split-fluorescent protein fused to an RNA binding domain[74,75]. Since only the scaffolded protein units are able to reconstitute the split chromophore and fluoresce, they can be easily distinguished from the unbound ones. Such an approach hence achieves lower background signals than systems where auto-fluorescent proteins are directly tagged onto RNA. As the repertoire of aptamer-RNA binding protein pairs is being extended through several *in vitro* methods like systematic evolution of ligands by exponential enrichment (SELEX)[76,77,78] newer combinations are being used to explore cellular function[79]. The studies discussed here have led to a better understanding of RNA diffusion and localization[32,74] in bacterial cells and measurement of transcriptional kinetics[80]. These efforts also enabled localization of a diverse array of proteins (such as enzymes) on RNA scaffolds, opening up applications in therapeutics and metabolic engineering.

## Packaging Therapeutics on RNA Scaffolds

RNA-RNA interactions can be used for exciting scaffold applications *in vivo*. Packaging RNA (pRNA) from bacteriophage, ϕ29 has been used as a building block for bottom-up assembly of drug delivery vehicles[18,81]. pRNA monomers consist of structural hairpin regions and dimerization/polymerization domains. Ends of the hairpin regions offer sites for tagging with drugs or targeting molecules. The polymerization domains can be engineered to favor formation of dimers, trimers, pentamers or hexamers as stable drug carriers[18,52,81]. Heterodimers containing pRNA tagged with a CD4 aptamer, and pRNA attached to an siRNA were shown to specifically target CD4-expressing T cells, leading to cell death[81]. This *in vitro* study also showed stability and efficacy of the nanoscale drug delivery particles for killing cancer cells. Such systems are advantageous since the pRNA polymers are hypothesized to be stable in physiological conditions, and be less immunogenic than protein carriers[81]. Finally, these polymers could be made specific to many *in situ* targets by using engineered specific RNA aptamers that recognize cellular moieties.

## Recombinant RNA Technology

RNA scaffolds have also been used to serve as protective tethers for the purification of recombinant RNA[82]. In this approach, a tRNA scaffold acts as a protective secondary structure to insulate the transcript from native *E. coli* nucleases and thereby stabilize production of recombinant RNA (recRNA) *in vivo*. The characteristic clover-leaf tRNA structure formed around a recRNA is recognized by native cellular enzymes and processed as tRNA. This ensures that each single transcript is a product of specific defined length. A sephadex affinity tag was included in the expressed sequence to allow

purification of transcripts that contained RNAs of medical research interest, like the human HBV-epsilon[82]. This design thus enables collection of large amounts of purified RNA transcripts for *in vitro* structural studies and vaccine development. Recently, these efforts have been extended to expression and purification of RNA-protein complexes[83], providing pure samples which could be used for crystallographic studies of natural RNA-protein interactions and potential use in cell-free systems.

## Localizing metabolic enzymes on RNA

As we have seen earlier, scaffolding and compartmentalization are effective strategies for optimization of metabolic pathway performance both in natural and synthetic systems[20,26]. A few studies have used DNA structures to coordinate the assembly of enzymes and study effects of spatial co-localization *in vitro*[27,84,29] and *in vivo*[31]. Protein scaffolds have also been used to channel metabolic substrates between co-localized enzymes in living cells[14,30]. Scaffolding is seen as a powerful tool to specifically direct metabolic pathway flux towards enzymes of choice, prevent loss of intermediates to competing reactions, and protect the host cell from any toxic or volatile intermediates through confinement at a sub-cellular location.

A notable effort in the use of RNA scaffolds for metabolic channeling achieved a nearly fifty-fold increase in hydrogen gas production in *E. coli*[13]. This effort combined many of the techniques discussed above. Synthetic RNA strands comprising of polymerization domains and aptamers for MS2 and PP7 coat proteins were expressed in the bacteria. Dimerization and polymerization domains allowed for tiling and assembly into a supramolecular structure. The large (40-100 nanometers) intra-cellular RNA assemblies greatly enhanced flux of electrons from ferrodoxin to hydrogenase when both enzymes were tethered to the scaffold with fusions to MS2 and PP7. Furthermore, sig-

nificant differences in titer were observed for scaffolding structures having different geometries, tying metabolic flux to the specific spatial positioning of the scaffold. Such an approach brings modular design and scalability[85] to metabolic engineering for biofuels and high value chemical synthesis, where control of intermediate metabolite flux can be critical[86,87,88].

## 1.3 Implementing control of metabolic flux

This thesis presents research efforts aimed at controlling metabolic fluxes in the bacteria *Escherichia coli* and *Synechococcus elongatus* to expand the synthetic biology toolkit available for metabolic engineers. *E. coli* has been a workhorse model system for studying and engineering bacterial metabolism[89]. *S. elongatus* is a photoautotrophic cyanobacterium and is also genetically tractable[90], making it a desirable platform for biological production without competing with traditional feedstocks[91]. The central strategy discussed here attempts to increase the yield of metabolic pathways by spatially co-localizing selected enzymes to enhance metabolite transfer rates.

In the next chapter we characterize and expand the RNA scaffold toolkit for *in vivo* use. We demonstrate that the aptamers and polymerization domains of scaffold strains can be treated as modular pieces. We characterize a library of aptamers and show that their corresponding RNA binding domains (RBDs) are able to reliably recruit split GFP fragments to scaffolds inside *E. coli*. This library includes aptamer-RBD pairs with a wide range of RBD sizes and dissociation constants, while being mutually orthogonal.

In chapter 3, we use these aptamers on RNA scaffolds to bring together two cyanobacterial enzymes that convert fatty acid synthesis intermediates to alkanes in *E. coli*. We

show that scaffolding leads to increased channeling of the intermediate fatty aldehyde resulting in higher alkane yields at the expense of side products. Interestingly, we find that changing the geometric orientation of enzymes on *in vivo* scaffolds leads to changes in intermediate flux and final alkane yields. We present a molecular scale model to explain this effect. In addition, we show that the scaffolds can also be used to tether 3-4 enzymes while channeling multiple chemically diverse metabolites.

In chapter 4, we turn to *S. elongatus*. In one approach, electronic flows in the cyanobacterium are manipulated for the biosynthesis of hydrogen gas[86], which is increasingly seen as a feasible alternative to fossil fuels[92]. Finally, we present efforts to express RNA scaffolds in *S. elongatus* and to use them as synthetic carbon fixation platforms.

Chapter 5 summarizes the results and offers a look towards the future of metabolic engineering and the role of RNA as a tool for synthetic biology.

*In conclusion, we suggest that the unit in structure A is, as in structure B, two co-axial helical chains running in opposite directions.*

Rosalind Franklin and Raymond Gosling, *Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate*[93]

# 2

# Expanding the toolkit for *in vivo* RNA scaffolds

The double helical structure[93] of nucleic acid strands with complementary base-paired[94] sequences allows for storage of hereditary information, translation of it into functional protein and RNA units and as noted in chapter 1, also enables nu-

cleic acids to play important structural roles in nature. Motivated by the aim of using tools offered by nucleic acid nanotechnology[95,13] for effective microbial metabolic engineering, we carried out the set of experiments presented in this and the next chapter.

Here, we* express synthetic RNA strands comprised of polymerization domains and aptamers, in bacterial cells[85]. The polymerization domains allow for assembly of RNA strands into a supramolecular structure[13]. We show targeted assembly onto such structures using fluorescent protein fusions to a library of aptamer-binding domain pairs that includes small RNA binding peptides[97] as well as natural domains like MS2 and PP7. We show that modular aptamers on RNA scaffolds of a variety of polymerizing dimensions show corresponding effects in co-localizing the proteins.

## 2.1 Expanding the repertoire of functional RNA binding domains

We have expanded the set of RNA binding domains that can be used to form fusion proteins that will bind to specific sites on our RNA scaffolds. The basic strategy builds on a scheme in which engineered *E. coli* express proteins fused to RNA binding domains (RBDs) that would assemble onto RNA scaffolds that include aptamers and polymerization domains. Strand design allows the aptamers to be presented on discrete (0-dimensional), 1-dimensional, or 2-dimensional scaffolds[13] (Fig 2.1). Each protein to be docked binds to its cognate binding site on the RNA scaffold. In previous work, RNA binding domains from viral coat proteins like MS2 and PP7 were fused with fluorescent proteins or enzymes for co-localization on RNA strands[13,32]. These viral RBDs bind tightly to their RNA aptamers z[98,45], but are relatively large (>120 amino acids) and hence may affect folding or activity of the resulting fusion protein through steric

*The work presented in this chapter was carried out with Abhishek Garg, and has been published[96]

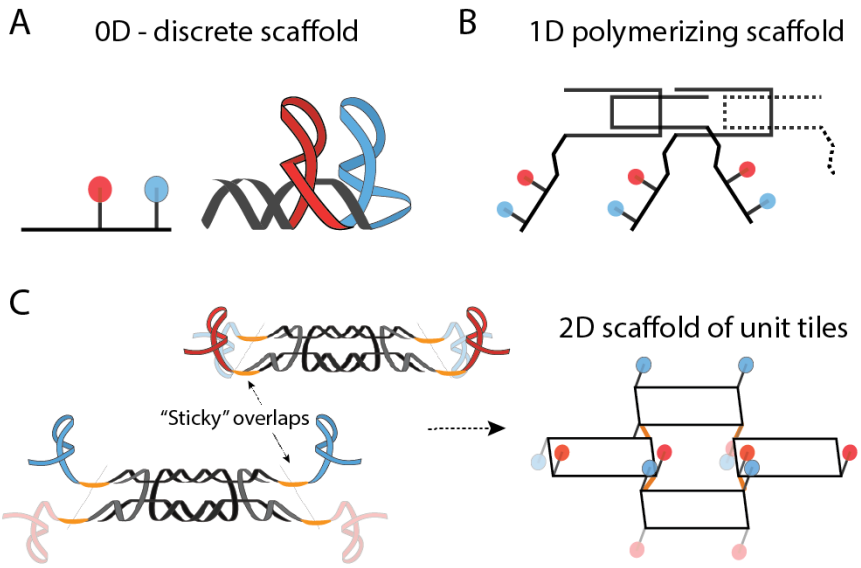



**Figure 2.1:** Three dimensionalities of RNA scaffolds. Our designs allow formation of three different types of RNA scaffolds [13]. (A) In the 0D case, a single RNA strand folds into a discrete unit presenting two aptamers. (B) The 1D polymerizing strands have 'sticky' ends that allow the individual units to form linear chains of aptamer sites through complementary base pairing. (C) In the case of 2D scaffolds, two different RNA oligonucleotides, A and B, come together to form unit tiles $A_2B_2$, which then polymerize in two dimensions through sticky end base pairs. Each corner of the unit-tile interactions brings together two different aptamers in close proximity. Red and blue represent different aptamers and dimmed shades show aptamers in a pointing downwards from the scaffold plane.


**Figure 2.2:** In our assay for the formation of aptamer binding site presenting RNA scaffolds, we co-express the RNA strand with two halves of a split GFP protein [99]. If the RNA scaffold folds into the designed secondary structure, the two halves are brought in close proximity through aptamer-RNA binding domain interactions and this complementation results in higher cellular fluorescence, which can be assayed by microscopy or flow cytometry.

interactions. Bayer et al [97] characterized several short arginine-rich motifs that bind RNA, along with their synthetically selected RNA aptamers. We chose 6 of these short RBD-aptamer sets (App. Fig A.1) along with MS2 and PP7, to create a library of 8 such interactions for *in vivo* applications. The 6 sets were picked so that we could obtain a set of binding domains that were both short (<25 amino acids), to enable facile tagging on enzymes, and also mutually orthogonal in binding to the corresponding aptamers [97] (App. Fig A.2).

To test *in vivo* RNA-protein interactions for the library obtained above, we made fusions of the RBDs in four pairs with inactive green fluorescent protein fragments (split GFP) A and B respectively (Fig 2.2 and Fig 2.3). The fragments were split as described by Valencia-Burton et al [74]. Corresponding aptamers for each pair of RBDs were included in RNA sequences, with their secondary structures designed to fold into discrete scaffolds (Fig 2.2). These '0D' scaffolds were co-expressed with split GFP-RBD fusions through plasmids in *E. coli*. In control cells, we expressed only the split GFP-RBD fusions with an empty vector so that no scaffolds for co-localizing the fragments were present. After 3 hours of induction, the cells were visualized at 100x magnification under a microscope to assay cellular fluorescence indicating co-localization of the split GFPs to create active GFP.

| Fusion | Aptamer | RBD | Kd (nM) |
|---|---|---|---|
| GFPA-MS2 | 14 bp | 131 aa | 2 |
| GFPB-PP7 | 25 bp | 124 aa | 2 |
| GFPA-RevN7D | 30 bp | 17 aa | 120 |
| GFPB-RevR11Q | 30 bp | 17 aa | 100 |
| GFPA-BMVGag | 30 bp | 19 aa | 20 |
| GFPB-lambdaN | 18 bp | 22 aa | 90 |
| GFPA-BIV-Tat | 28 bp | 17 aa | 60 |
| GFPB-HTLV | 36 bp | 16 aa | 270 |

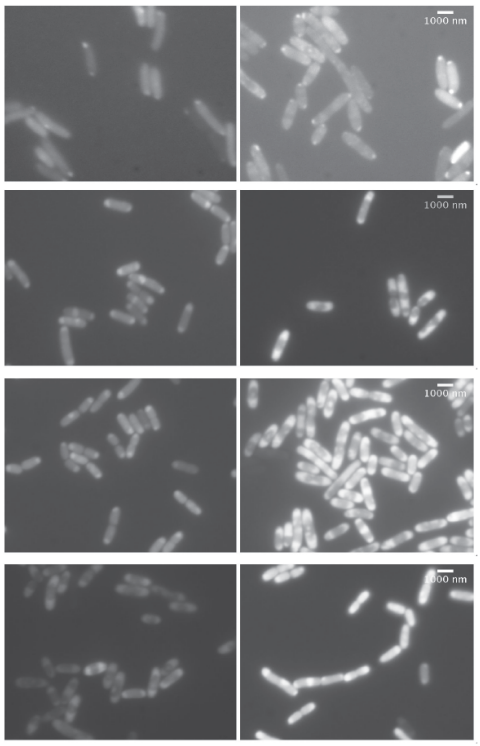



**Figure 2.3:** This figure shows a list of 4 different pairs with 8 unique aptamer-RNA binding sets tested. Values of dissociation constants reported previously [97,98,45] are tabulated with domain and aptamer sizes. Also shown are FITC images revealing enhanced split GFP fluorescence *in vivo* in the presence of RNA scaffolds for all sets tested. Scale bars: 1 micrometer.

| Fusion | Aptamer | RBD | Kd (nM) |
|---|---|---|---|
| GFPA-MS2 | 14 bp | 131 aa | 2 |
| GFPB-PP7 | 25 bp | 124 aa | 2 |
| GFPA-RevN7D | 30 bp | 17 aa | 120 |
| GFPB-RevR11Q | 30 bp | 17 aa | 100 |
| GFPA-BMVGag | 30 bp | 19 aa | 20 |
| GFPB-lambdaN | 18 bp | 22 aa | 90 |
| GFPA-BIV-Tat | 28 bp | 17 aa | 60 |
| GFPB-HTLV | 36 bp | 16 aa | 270 |

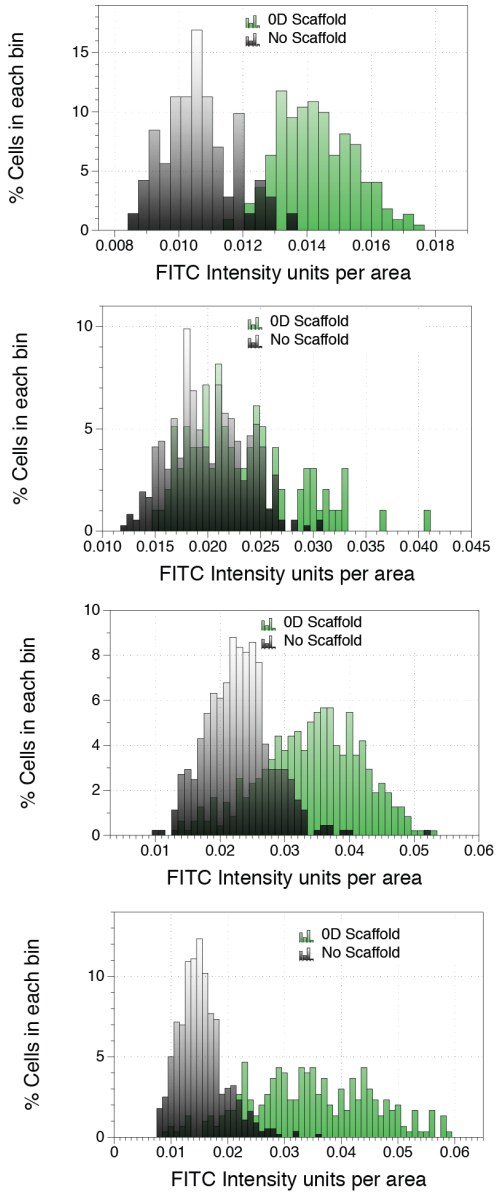



**Figure 2.4:** Histograms of FITC intensity profiles for 4 pairs of split GFP - RBD fusions tested on 0D scaffolds in Fig 2.3. Differences in FITC intensity scales depended on the FITC exposure times chosen (50ms - 300ms) to ensure signal profiles were not saturated.

Protein-RNA binding *in vivo* was in evidence for all 8 aptamer-RBD sets tested in 4 pairs (Fig 2.3). When the split GFP fragments are expressed without scaffolds the cells are only dimly fluorescent; upon co-expression with the scaffold sequence, however, significantly brighter cells are observed, presumably because the split GFP fragments bind adjacent to each other on the RNA scaffold (Fig 2.3). Histograms showing average cellular FITC intensity from 100-400 cells for the 4 pairs are presented in Fig 2.4. For all four pairs tested, the histograms show a clearly brighter population of cells (right-shifted distribution) when the scaffold RNA was co-expressed with the split GFP-RBD fusions (Fig 2.4). We have previously shown such split GFP complementation on RNA scaffolds for the PP7 and MS2 domains' pair[13], which served as a positive control here. We also observed enhanced GFP complementation on scaffolds from three pairs of the 6 new aptamer-RBD partners. This gives us a tool-kit of eight potential aptamer-protein interaction sets with varying sizes and binding properties.

## 2.2 RNA binding domains function on polymerizing RNA scaffolds

We showed that the split GFP complementation is efficient on RNA scaffolds of higher dimensions (1D and 2D) with the new RBDs and aptamers as well. For this we chose two aptamer-RBD interaction sets: PP7 and BIV-Tat. We chose PP7 because of its tight binding and ability to form non-aggregating dimers on RNA aptamers[45]. BIV-Tat was chosen because it was the strongest-interacting short domain for which a crystal structure of the aptamer-RBD complex is known[97,100] (App. Fig A.2). As above, we expressed fused split GFP fragments with a plasmid expressing either no RNA or RNA strands for 0D, 1D, or 2D RNA scaffolds (Fig 2.5A). Fluorescence microscope images of the cells were analyzed using software packages MicrobeTracker[101] and ImageJ[102] to

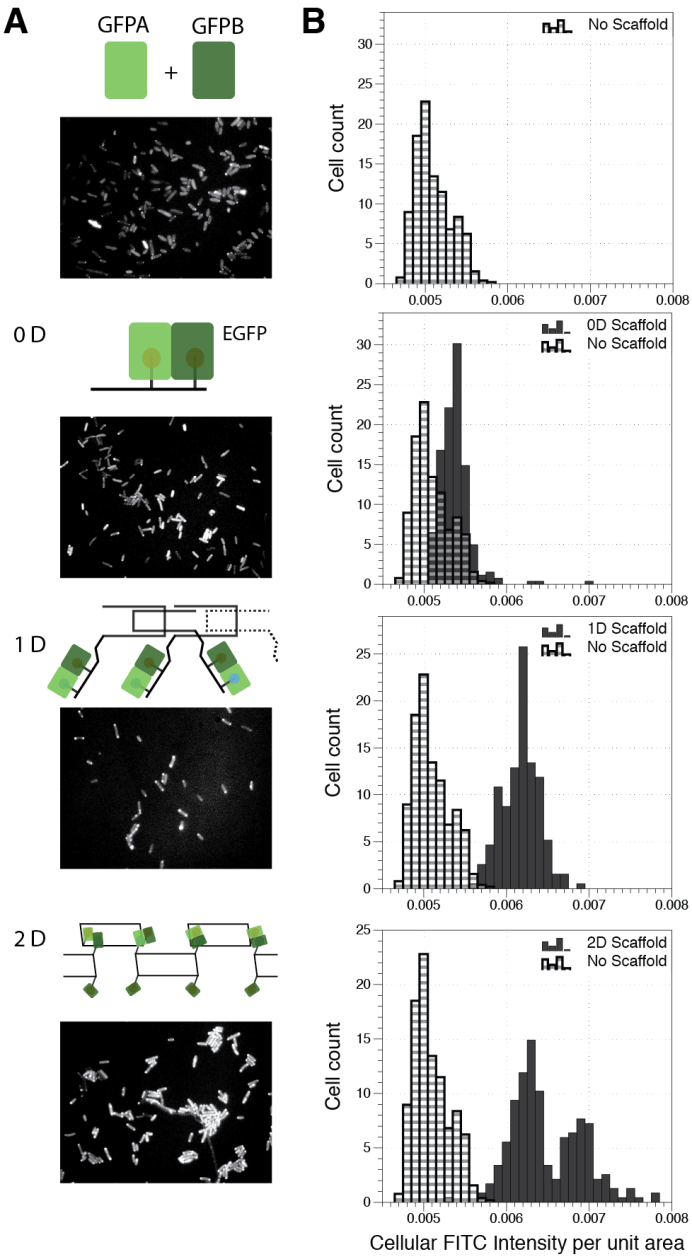



**Figure 2.5:** Aptamers incorporated into 0, 1, and 2 dimensional RNA scaffolds bind proteins *in vivo* (A) FITC images showing GFP fluorescence observed in cells on expression of split GFP fragments (GFPA-PP7 and GFPB-BIV-TAT) with either no scaffold, 0D, 1D or 2D scaffolds containing PP7 and BIV-TAT binding aptamers (B) Corresponding FITC quantification (histograms) for 0, 1 and 2D scaffold designs. ($p < 10^{-3}$, one-tailed t-test)

quantify average fluorescence per unit area for several images (>100 cells) of each strain. We performed statistical analyses (Fig 2.5B) to calculate the average cellular fluorescence intensity in each case. A progressive shift towards brighter cells upon expression of RNA scaffolds of the three dimensionalities was observed (Fig 2.5A,B). The automated analysis indicated two apparent populations when the GFP fragments are co-expressed with the 2D scaffold; visual inspection of the fields confirmed that there are two populations of bright cells. We hypothesize that the kinetics of RNA tiling lead to extended 2D scaffold assembly in the brightest population of cells either due to small differences in RNA expression levels or stochastic coalescing of multiple assemblies.

## 2.3 Discussion

We show *in vivo* functionality of a library of modular RNA binding domains and aptamer scaffolds. These RBD-aptamer sets are effective in co-localizing proteins within living *E. coli*. The RBD-aptamer sets were used to reconstitute split GFP fragments on a variety of RNA scaffolds; 0, 1, and 2 dimensional RNA scaffolds, showed progressively more noticeable concentration of proteins. Split GFP fragments offer a handy tool for probing supramolecular assemblies where two disparate units needs to be spatially united. However, imaging of GFP foci often needs co-localization of tens of proteins[32] and due to diffraction limits of optical imaging, precise geometric details of the assemblies are hard to capture. The modularity of our scaffolds would allow tagging with oligonucleotides that enable super-resolution imaging[103]. Further, use of non-protein fluorophores[104] might help better understand the kinetics of scaffold assembly itself.

The library of 8 RNA binding domain-aptamer sets tested here forms a useful

toolkit for designing in-vivo protein-RNA interactions. The library of protein/RNA pairs is orthogonal and includes a range of aptamer lengths, binding domain sizes and binding dissociation constants (App. Figs. A.1, A.2). Characterization of shorter RNA binding peptides will, in principle, enable facile engineering of enzyme fusions with minimal perturbation to protein structure. We have also demonstrated the modular nature of our RNA scaffold design, by using the aptamer site library and showing assembly of functional discrete, 1D or 2D scaffolds. In the next chapter we employ these scaffolds and aptamer-RNA binding pairs to show that enzymatic pathways can be spatially co-localized *in vivo* to achieve higher metabolic fluxes in a geometrically dependent manner.

## 2.4 Materials and Methods

### 2.4.1 Strain construction

*E. coli* strain BL21* (DE3), that carries a mutation is the me131 (RNaseE) gene, was used for all experiments described here. The appendix Fig A.3 lists the plasmid systems used for expression of the various components in *E. coli*. RNA and protein expression was based on Novagen (EMD Millipore) Duet vectors. pETDuet1 was used for RNA expression because of its high copy number (>40 copies/cell) while the lower copy pACYCDuet1 (10-15 copies/cell) and pCOLADuet1 (20-40 copies/cell) were used to express proteins. For co-expression, we co-transformed competent BL21* DE3 cells with the two or three relevant plasmids and selected on plates with a combination of antibiotics.

### 2.4.2 Microscopy and analysis

For GFP microscopy experiments, cells were grown in LB media and induced with 0.3 mM IPTG at OD 0.8. Cells were allowed to grow for 3 more hours at 37°C with shaking. They were then washed with M9 media to reduce the background fluorescence from LB media. Cells were placed on 2% clear agarose pads and analyzed using a Nikon TE-2000 microscope (100x, 1.4 numerical aperture objective, ORCA-ER charge-coupled device camera, FITC channel. The images were analyzed using the software packages ImageJ and MicrobeTracker[101,102]. We calculated the average cellular fluorescence using MicrobeTracker and plotted the distribution as histograms using the software DataGraph 3.1.1. 2013. Visual Data Tools, Inc.

*At the atomic level, we have new kinds of forces and new kinds of possibilities, new kinds of effects... I am, as I said, inspired by the biological phenomena in which chemical forces are used in a repetitious fashion to produce all kinds of weird effects.*

Richard P. Feynman, *There's plenty of Room at the Bottom*[105]

# 3

# Increase in metabolic output from RNA scaffolds

Biomolecular organizations achieve nanoscale precision leading to the complex properties and behavior that Feynman[105] marveled that. Synthetic biology today enables the kind of molecular level engineering that he presciently suggested.

Continuing with the aim of using the precise molecular interactions for increasing the flux of selected metabolic pathways, this chapter details interesting novel effects at the nanoscale that facilitate such applications.

The role of enzyme co-localization in pathways that involve diffusible intermediates continues to be debated[106]. While some analyses rely on three dimensional diffusion models to argue that the effectiveness of substrate channeling may be limited to cases where intermediate pools are dilute[33], others have suggested a more general 'molecular crowding' hypothesis[12]. The latter view is supported by a recent study, in which the authors demonstrated increased transfer of hydrogen peroxide between enzymes scaffolded in vitro[28]. In that study, a discontinuous increase in intermediate channeling was observed for proteins tethered less than 20 nm apart. The authors proposed a mechanism of metabolite channeling by restricted diffusion on hydration layers across crowded protein surfaces. Metabolite diffusion may be particularly restricted for hydrophobic substrates such as fatty acids and related molecules, which may associate with protein surfaces more strongly than hydrophilic substrates. In this regard, the two-step alkane synthesis pathway described by Schirmer et al.[87] is of interest because it involves an uncharged aldehyde intermediate. The two enzymes, acyl-ACP reductase and aldehyde deformylating oxygenase[107] can be heterologously expressed in *E. coli* to produce pentadecane. Also of interest is the channeling of cofactors and carboxylic acids, since they form intermediates in several industrially relevant metabolic products. Succinate is one such product where pathway engineering has focused on increasing supply of oxaloacetate[108], and NADH[109]. Co-localization of enzymes such that the fatty aldehyde intermediate, or oxaloacetate and NADH are transferred directly from the surface of one enzyme to the other could enhance production of pentadecane and succinate respectively.

Here, as in chapter 2, we* express synthetic RNA strands comprised of polymerization domains and aptamers, in *E. coli*. We have seen that polymerization domains allow for assembly of RNA strands into a supramolecular structure upon which proteins can be effectively co-localized. In experiments described in the current chapter, fusions of RNA-binding domains with enzymes involved a two-step pentadecane pathway are co-expressed to allow assembly onto the RNA scaffolds. We show that increased production of pentadecane depends on the relative geometric orientation of enzymes. Further, we report higher metabolic output on RNA scaffolds with up to 4 enzymes co-localized from a multi-step succinate production pathway. Our results suggest that the specific orientation and placement of enzymes on the scaffold enhances pathway flux consistent with restricted substrate diffusion.

## 3.1 Increased pentadecane production on RNA scaffolds

To use RNA scaffolds for metabolic substrate channeling, we employed a two-step pentadecane production pathway (Fig 3.1) which has recently been characterized using cyanobacterial enzymes expressed in *E. coli*[87]. Two enzymes, acyl-ACP reductase (AAR), and aldehyde deformylating oxygenase (ADO), sequentially convert cellular N-carbon fatty acyl-ACPs to alkanes of length N-1. For example, AAR reduces the C16 acyl-ACP molecule to produce a fatty aldehyde, hexadecanal ($C_{15}H_{31}CHO$), which undergoes deformylation catalyzed by ADO[107] to form pentadecane. However, upon co-expression of these two enzymes in *E. coli*, the major products are alcohols and not alkanes[87]. The intermediate aldehyde appears to be be non-specifically reduced by aldehyde reductases in the cell[110]. We hypothesized that if hexadecanal could be selectively

*The work presented in this chapter was carried out with David Godding and Jeffrey Way, and has been published[96]

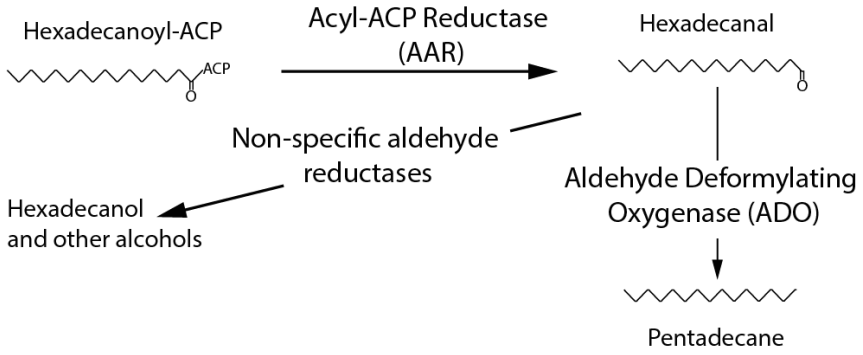



**Figure 3.1:** The pentadecane production pathway described by Schirmer et al [87] composed of two cyanobacterial enzymes Acyl ACP Reductase (AAR) and Aldehyde Deformylating Oxygenase (ADO). Also shown is the competing side reaction that leads to alcohol production, including hexadecanol.

channeled from AAR to ADO on scaffolds, an increase in pentadecane would occur.

Genetic fusions of AAR and ADO to RNA binding domains were constructed on the N-termini of the enzymes. AAR was tagged with BIV-Tat and ADO was fused to PP7; both fusions retained activity (Fig 3.2), although attempted fusions with the MS2 failed to show activity. *E. coli* expressing the two enzyme fusions were grown and alkane levels in the supernatant were measured (see section 3.5 - Materials and methods). Pentadecane production, which is absent in WT *E. coli*, was observed (Fig 3.2) confirming catalytic activity of the fusion proteins. We measured alkane production after 16 hours of log-phase growth in batch cultures, while earlier work has reported higher levels can be achieved from up to 40 hours of induced growth [87]. In addition, we also observed other alkane products however, levels were much smaller than pentadecane, consistent with previous reports [87], and could not be accurately quantified.

When the enzyme fusions were co-expressed with 1D, or 2D RNA scaffolds containing their cognate binding sites (Fig 3.2) in *E. coli*, alkane production was enhanced by 80% compared to co-expression with an empty vector or a 0D scaffold, as assayed

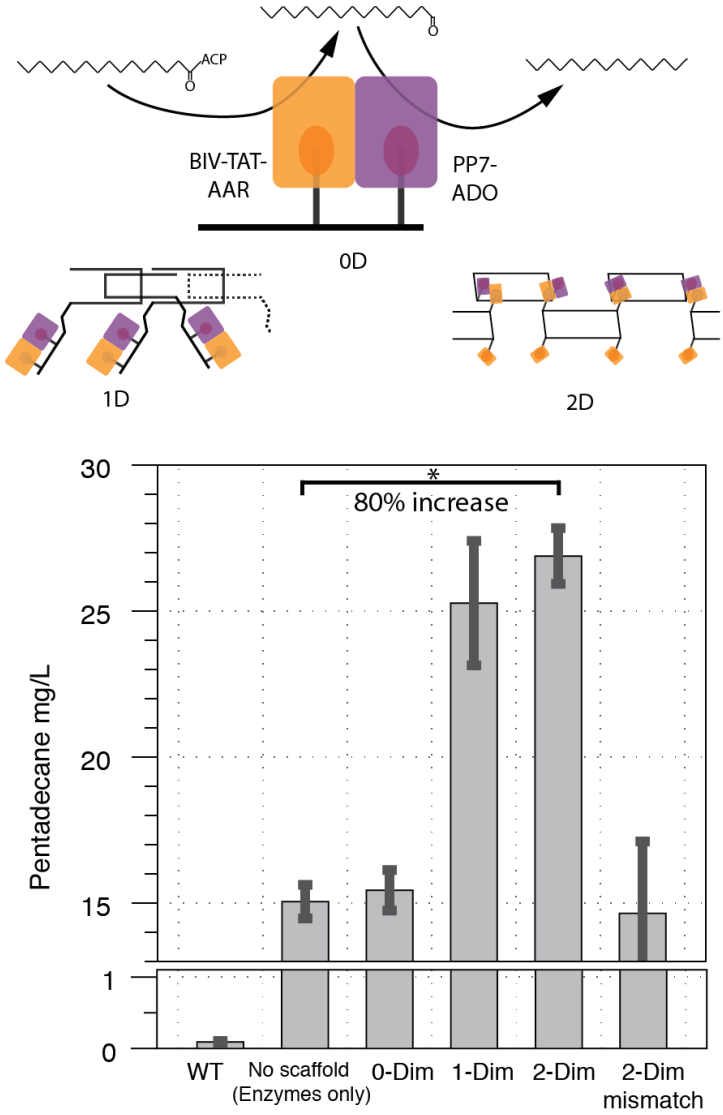



**Figure 3.2:** This figure shows the scaffolding design implemented to channel the hexadecimal intermediate towards pentadecane production. Fusions of AAR and ADO with BIV-TAT and PP7 respectively are co-expressed in *E. coli* with scaffolds of different dimensions carrying the corresponding binding sites. On assaying different levels of pentadecane are observed with either an empty vector (no scaffolds) or RNA scaffolds or 0,1, or 2 dimensions (with aptamers for BIV-TAT and PP7), or a 2D scaffold with mismatched aptamers (containing anti-RevR11Q and MS2). (n=3, error bars = SEM) (* indicates $p < 0.05$, one-tailed t-test)

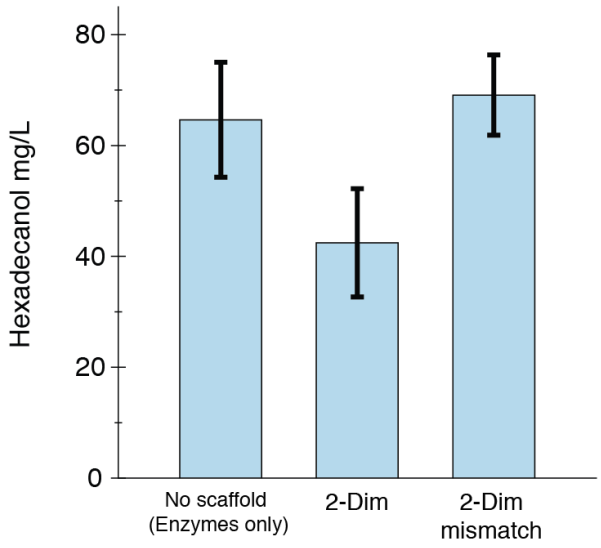



**Figure 3.3:** One of the pentadecane pathway side-products, hexadecanol, production measured on expression of pathway enzymes with the same empty vector, correct 2D scaffold, and mismatched 2D scaffold. (n=3, error bars = SEM)

by pentadecane levels in ethyl acetate extracts (Fig 3.2). We also measured production of hexadecanol, one of the alcohol side products reported when AAR and ADO are co-expressed[87]. We observe that the hexadecanol levels are reduced when enzymes are co-localized on their cognate 2D scaffolds, compared to expression without any RNA scaffolds (Fig 3.3). As a control, expression of the enzymes with a 2D scaffold lacking the cognate aptamers showed no difference from the empty vector control (Figs 3.2, and 3.3). The mismatched scaffold carried binding sites for RevR11Q and MS2 instead of the BIV-TAT and PP7 aptamers. Lack of any effect on the mismatched scaffold shows that co-expression with the correct cognate scaffold was necessary for an effect.

We also sought direct evidence of enzyme-RNA scaffold interaction in our co-expression system. For this, we UV-crosslinked protein-RNA interactions in cells expressing the pentadecane pathway enzymes and co-purified BIV-TAT-AAR with any

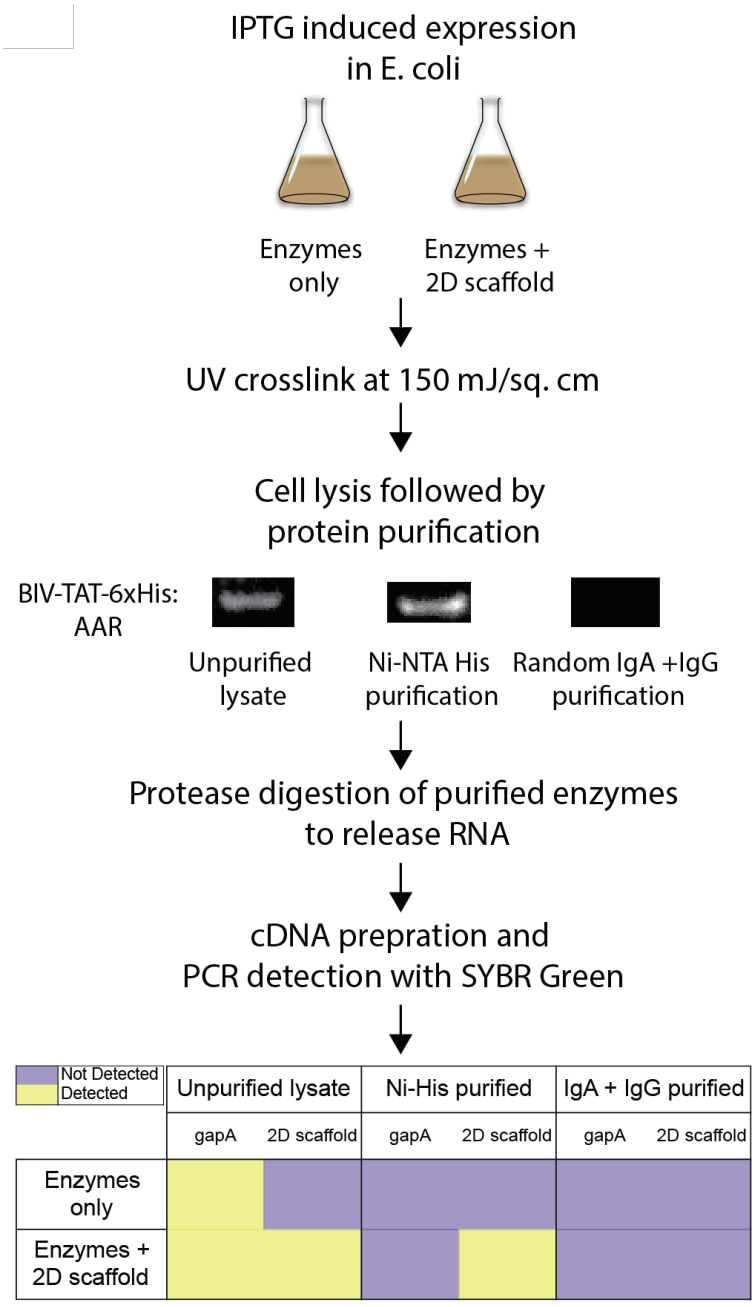



**Figure 3.4:** Direct evidence of enzyme-RNA interaction from experimental flow for detecting RNA associated in-vivo with 6xHis tagged BIV-TAT-AAR enzymes. Purple boxes stand for amplification detected by SYBR green and yellow boxes indicate that the RNA species was below detection threshold.

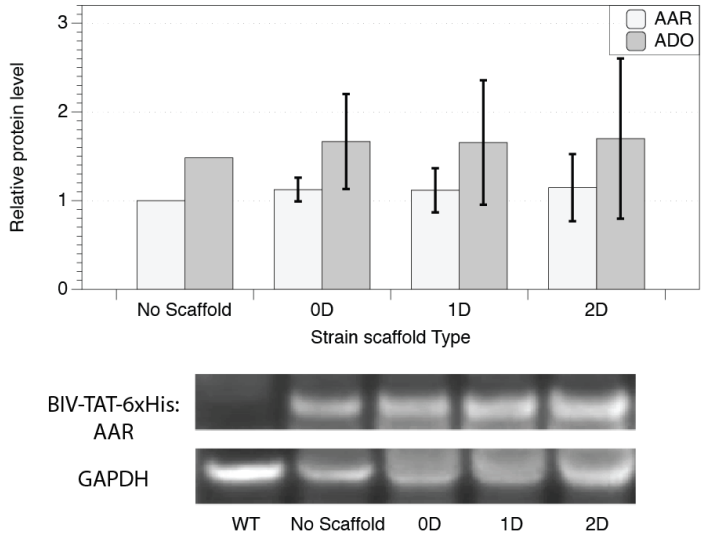



**Figure 3.5:** Pathway enzyme production levels measured by quantitative western blotting using 6xHis (AAR) and Strep (ADO) tags and normalized to GAPDH. Levels are shown for co-expression with an empty vector or scaffolds of different dimensionalities. (n=3, error bars = SEM)

bound RNA (Fig 3.4). BIV-TAT-AAR was tagged with 6xHistidine and could be seen only upon purification with Ni-NTA magnetic beads and not when random IgA + IgG beads were used as control (Fig 3.4). After digestion of the protein and RNA purification, we looked for presence of scaffold RNA levels by quantitative RT-PCR. Internal primers were designed to amplify scaffold RNA and as well cellular gapA mRNA. While scaffold RNA and gapA are both detected in unpurified cell lysates expressing 2D scaffolds, only the scaffold RNA is detected from the extracts co-purified with BIV-TAT-AAR. Neither gapA, nor scaffold RNA are detected from the IgA+IgG pull downs. This suggests that while a control mRNA like gapA doesn't interact with our enzymes appreciably, interactions of BIV-TAT-AAR and the co-expressed 2D scaffold are significantly higher. To ensure that the enhanced synthesis reflected in Fig 3.2 was due to scaffolding of the enzymes and not differential expression, we compared

protein levels using Strep and 6x His tags included in the fusions for AAR and ADO respectively. Quantitative Western blots were prepared on cell extracts and enzyme levels were found to be unaffected by co-expression with scaffolds (Fig 3.5). This indicates that when we expressed the two-enzyme pathway in the presence of 1 or 2 dimensional polymerizing RNA scaffolds, higher pentadecane titers were realized for the same amounts of proteins expressed.

### 3.1.1 Alkane yield enhancement highest in exponential growth phase

In applied and industrial settings continuous cultures are more commonly employed since they allow for faster process rates, lesser down time, and better controlled standard conditions[111]. One significant difference between batch and continuous modes of growth is the variability in growth rate. While continuous cultures maintain uniform growth rates at all times, batch cultures follow sigmoid growth curves, with the exponential phase representing a sub-period of sustained uniform growth.

We tested the 2D scaffold producing strain for pentadecane yields at different points in batch culture and compared that to the strain carrying an empty vector (Fig 3.6). We measured pentadecane production from batch cultures at three different time points 16 hours (mid exponential phase), 20 hours (early stationary phase) and 26 hours (late stationary phase) after induction. As can be seen in Fig 3.6A, in general higher titers are observed for later time points. However, when we compare the fractional increase in pentadecane production due to scaffolding, the trend is reversed. At the 16 hours time point, we observe an 80% higher titer from 2D scaffolds. However, the fractional increase became smaller at 20 hours and less significant at 26 hours after induction (Fig 3.6A). Enzyme levels were again quantified by western blots and relative protein expression levels at the three time points were compared; the average enzyme levels

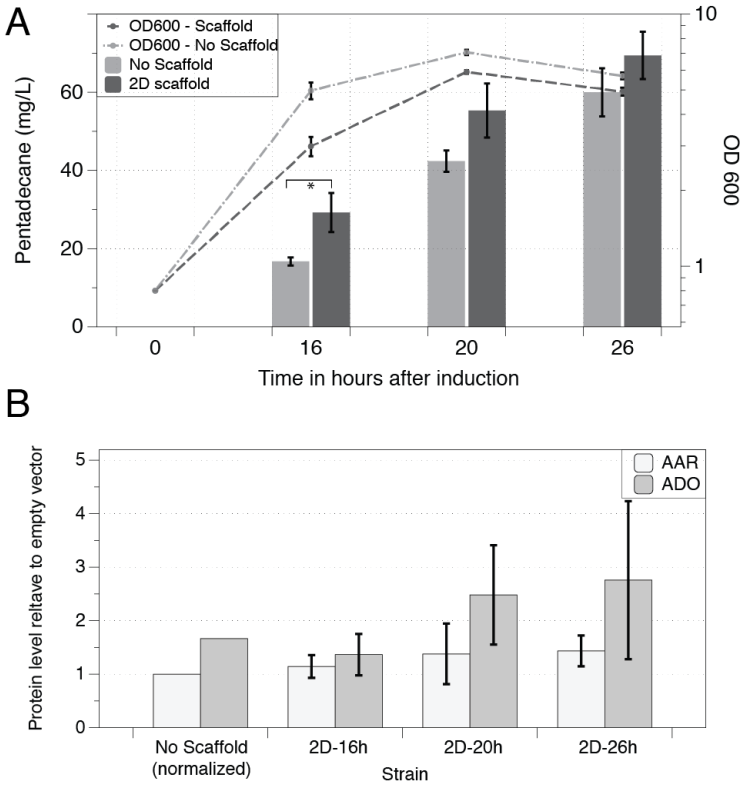



**Figure 3.6:** Pentadecane yield enhancement is highest in exponential growth phase (A) Time series of enhanced alkane production for 2D scaffolds at different time points in batch culture, plotted with cellular growth measured as optical density (OD) at 600 nm. (B) Enzyme levels measured by quantitative western blotting using His (AAR) and Strep (ADO) tags and normalized to GAPDH. (n=3, error bars = SEM) (* indicates $p < 0.05$, one-tailed t-test)

were not significantly different at any of the time points (Fig 3.6B). However, the later measurements carried larger error bars, showing a marked increase in variability of enzyme concentration.

## 3.2 Flux enhancement depends on geometry of scaffolded enzymes

The RNA scaffold offers structural flexibility that might be further modified to enhance the interplay of associated enzymes. Some studies have probed the effect of varying numbers of binding sites[14] or different inter-enzyme distances on metabolic channeling[28]. Here, we investigate the effect of altering relative enzyme orientation on the scaffold. The structure of two-dimensional sheets (Fig 2.1) resulting from assembly of the RNA oligonucleotides (App. Fig A.4) that make up the scaffold can be inferred based on the helicity of double-stranded RNA (11 bp/turn), the structure of Holliday junctions[112], and general principles of nucleic acid nanotechnology[13,95]. Four strands ($A_2B_2$) come together to form a unit tile, which polymerizes in two dimensions through base pair interactions on the tile corners. The aptamers are presented on two different sides of the plane of the unit tile, as shown (Figs 2.1, and 3.7A). Each unit tile has four 6-bp sticky ends that base-pair with diagonally related units. Through these interactions, one PP7 aptamer and one BIV-Tat aptamer emerge on the same side of the scaffold plane at each corner of the unit tiles (Figs 2.1, and 3.7A).

We constructed a high-level model of this aptamer pair from PDB files of aptamer-binding protein complexes, double-stranded RNAs, and the deformylating oxygenase enzyme from a related organism[113,114,115]. The enzymes being co-localized on the scaffold bind and form PP7-aptamer (PDB 2QUX[113]) and BIV-Tat-aptamer (PDB 2A9X[114]) complexes. Such complexes can be treated as rigid elements composed of PP7-linker-

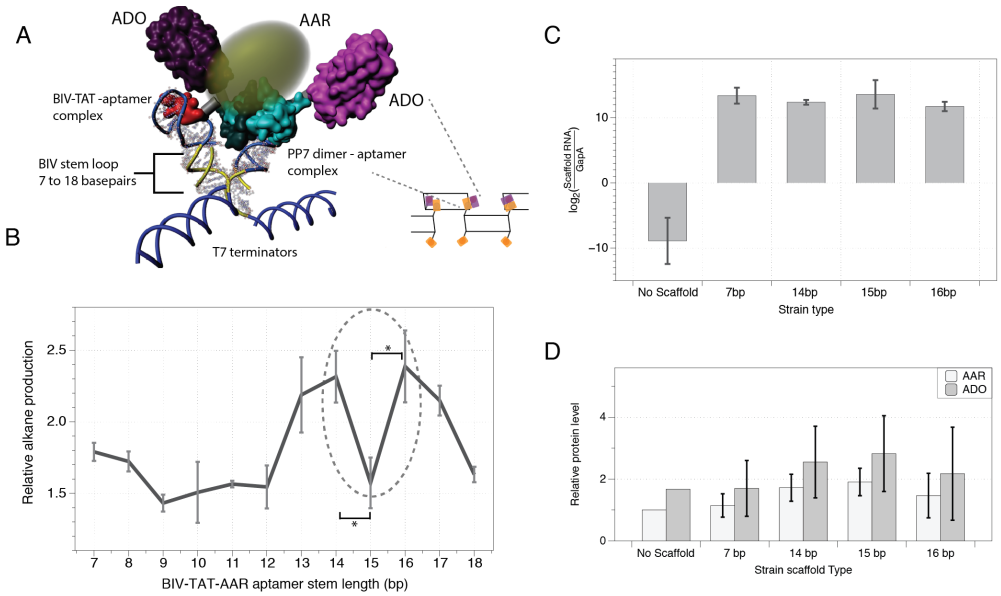



**Figure 3.7:** Length and orientation of aptamers affect metabolic flux increase on scaffolds. (A) Schematic showing design details of a pair of aptamers interacting on the 2D scaffold using PDB files, 2QUX [113], 2A9X [114], and 4KVQ [115]. Different base-pair lengths (7-18bp) were used for the hairpin presenting aptamer bound by BIV-Tat-AAR so that rotation of the complex may bring AAR in closer proximity to ADO on the scaffold (B) Pentadecane yields are shown for 2D scaffolds with the different BIV-Tat binding aptamer stem lengths. Concentrations were normalized to yield from enzyme expression without scaffolds and plotted as relative production levels. (C) RNA scaffold production levels, measured by qRT-PCR and normalized to endogenous gapA mRNA, for the strains showing greatest variation in yields. (Log base value $2r \approx 1.87$) see Materials and Methods (D) Pathway enzyme production levels measured by quantitative western blotting using 6xHis (AAR) and Strep (ADO) tags. Levels are shown for co-expression of pathway enzymes with RNA scaffolds of varying BIV-TAT-aptamer lengths and are normalized to GAPDH levels. (n=3, error bars = SEM) (* indicates $p < 0.05$, one-tailed t-test)

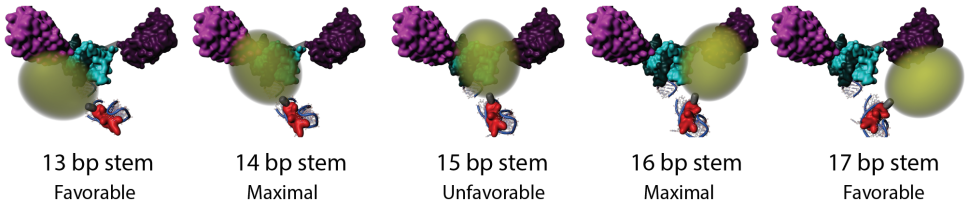



**Figure 3.8:** Proposed model for two maximal configurations of intermediate flux channeling. On varying the anti-BIV-TAT aptamer stem loop length, different rotational conformations of the BIV-Tat-AAR moiety are possible, relative to the PP7-ADO dimer.

ADO and BIV-Tat-linker-AAR fusion proteins which are expected to be under constrained Brownian motion as predicted by the coarse-grained molecular dynamics simulation approach of Robinson-Mosher et al[116]. In Fig 3.7A, the crystal structure of ADO is represented by Prochlorococcus marinus aldehyde-deformylating oxygenase (PDB 4KVQ[115]) which has 95% sequence identity to the S. elongatus enzyme used here. We found that systematic addition of base pairs to the stem-loop structure of anti-BIV-Tat aptamer identified scaffold variants with increased activity. To explore the effect of changing the position of the reductase, we added 0-11 base pairs to the BIV-Tat hairpin on the 2D scaffold, to go through an entire RNA helical turn (Fig 3.7A). We then measured the pentadecane synthesis from cells expressing the twelve 2D scaffolds with the BIV-Tat-AAR and PP7-ADO fusion proteins (Fig 3.7B). As before, the strains were grown aerobically, induced, and finally extracted with ethyl acetate for assaying pentadecane production. The enzymes co-expressed with an empty vector served as control. All designs showed increased pentadecane production relative to the no-scaffold case (Fig 3.7B). The increase in production ranges from 40% (9bp) to 140% (16bp) greater than co-expression without scaffolds.

A striking pattern of pentadecane synthesis was observed as a function of hairpin stem length, particularly with stems of 13-17 base pairs. Synthesis was near-maximal

with 13 and 17-bp stems, maximal with 14 and 16-bp stems, and minimal with a 15-bp stem (Fig 3.7B). These results may be interpreted in terms of a model (Fig 3.8) where the peaks in activity with stems of lengths 13, 14, 16, and 17bp may correspond to the acyl-ACP reductase being in proximity to one or the other of the deformylating oxygenases fused to the dimeric PP7 protein.

To ensure that these effects were solely due to changes in scaffold design, we measured the relative levels of scaffold RNA and enzyme production for strains with the 7, 14, 15, and 16 base pair scaffolds. We extracted total cellular RNA from cell pellets for each strain after an induction experiment and measured scaffold RNA levels by quantitative RT-PCR. Once again, primers to measure the 2D scaffold RNA level and endogenous cellular pools of gapA mRNA were used (Fig 3.7C). All measured scaffold RNAs were produced at about the same levels, with a variation of only 2 to 4-fold, corresponding to the error of qPCR. Enzyme levels were measured by quantitative western blots on cell pellet extracts from these strains (Fig 3.7D); no significant difference in enzyme concentrations between the strains and the no-scaffold controls were seen.

## 3.3 Localizing multiple enzymes on RNA scaffolds

RNA scaffolds can be used to enhance pathway flux in longer multi-step enzymatic pathways. We chose the commercially relevant succinate production pathway that has multiple rate limiting control points at the pyruvate utilization node[88] (Fig 3.9). Production of oxaloacetate through pyruvate carboxylase (PYC) enhances flux towards succinate[108]. Further, increased succinate production is noted when a $NAD^+$-dependent formate dehydrogenase (FDH) generates NADH, a co-factor used in the

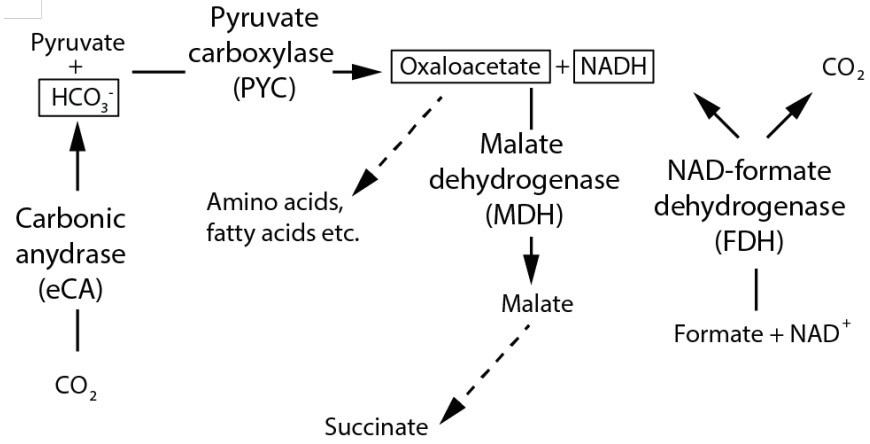



**Figure 3.9:** Conversion of pyruvate and oxaloacetate from significant control points for the fermentative production succinate from *E. coli*[88]. Shown here are the enzymes we expressed to increase succinate production in *E. coli*. We hypothesized that the intermediate molecules shown in boxes would be preferentially channeled towards succinate production.

conversion of oxaloacetate to malate, by malate dehydrogenase (MDH)[109].

Thus, we wanted to test if scaffolding of PYC (from *Rhizobium etli*), MDH (from *E. coli*), and FDH (from *Candida boidinii*) on a 2D scaffold results in increased redirection of pyruvate towards succinate in *E. coli* (Fig 3.10). We also tested a 4 enzyme co-localization scheme, in which a cyanobacterial carbonic anhydrase (eCA), known to increase oxaloacetate flux towards succinate[117] was included (Fig 3.10). We made fusions of PYC, MDH, FDH, and eCA with RNA binding domains PP7, BIV-TAT, lambdaN, and RevR11Q respectively. We observe that co-expression of PP7-PYC, BIV-TAT-MDH, and lambdaN-FDH with a 2D scaffold containing the corresponding aptamers leads to 83% higher succinate production than expression of the enzyme fusions with an empty vector (Fig 3.10). When RevR11Q-eCA is also expressed with these three enzymes, the new corresponding scaffold with 4 aptamers leads to an 88% increase (Fig 3.10). These results indicate that intermediates as diverse in their chemical composition

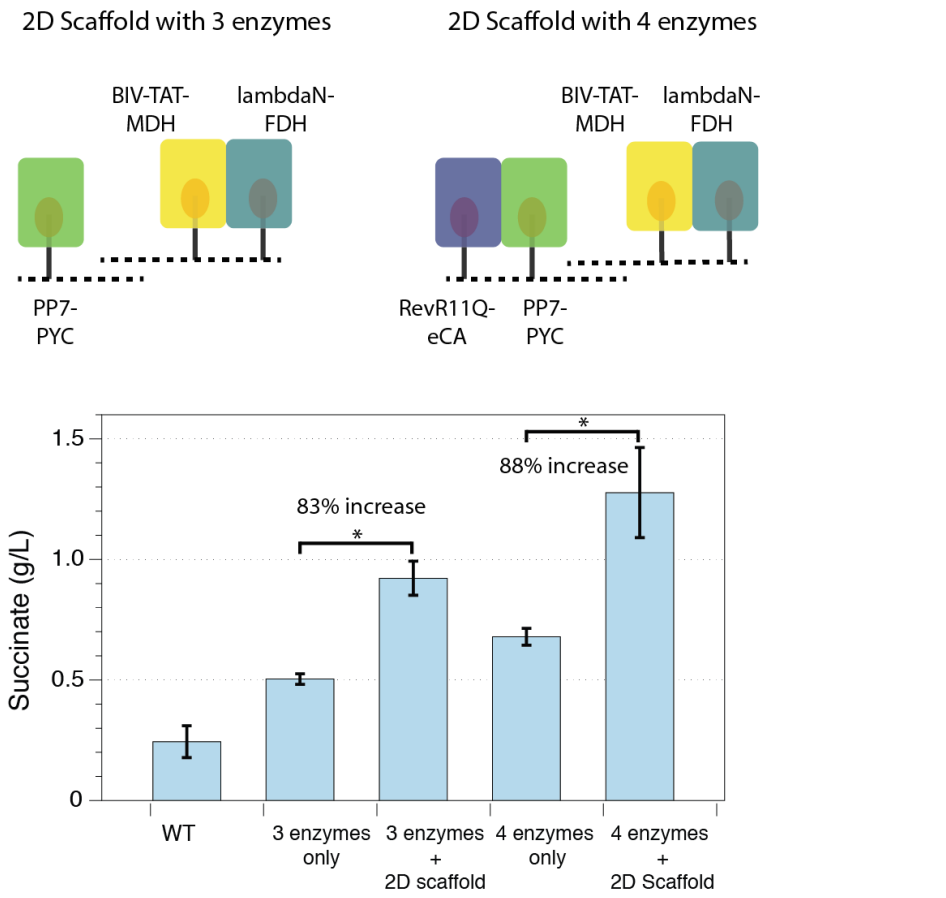



**Figure 3.10:** Schematics for two scaffolding approaches on 2D scaffolds to increase pathway flux towards succinate. PYC and FDH are scaffolded with MDH in the first approach to prevent loss of oxaloacetate and NADH respectively. In the scaffold with 4 aptamers, eCA is also scaffolded to provide $HCO3^-$ for PYC. The plot shows succinate production from *E. coli* expressing 3 enzyme fusions (PP7-PYC, BIV-TAT-MDH, lambaN-FDH) or 4 enzymes (also RevR11Q-eCA) with and without the corresponding 2D scaffolds compared to wildtype (WT) levels. (n=3, error bars = SEM) (* indicates $p < 0.05$, one-tailed t-test)

as oxaloacetate, NADH, and $HCO_3^-$ can likely be channeled towards desired product formation on RNA scaffolds with 2-4 different modular aptamer sites.

## 3.4 Discussion

Compartmentalization and scaffolding of enzymes are effective strategies for the control of cellular metabolism in both natural [20,21,22,23,24,25] and synthetic [20,13,14,26,27,28,29,30,31] systems. In this work we established a toolkit for designing protein-RNA interactions that work inside cells and used it to enhance metabolic substrate channeling in a biofuel production pathway. The key features of our system are the expression in cells of RNAs that form a multidimensional nanostructure scaffold with binding sites for RNA-binding domains (aptamers) and co-expression of fusions between specific RNA-binding domains (RBDs) and enzymes defining a metabolic pathway.

Having shown (in chapter 2) assembly of 0, 1, and 2 dimensional RNA scaffolds, which were progressively more effective in sub-cellular co-localization of proteins, we next used these tools to demonstrate metabolic channeling on RNA scaffolds inside cells. A two-enzyme pentadecane synthesis pathway with a labile aldehyde intermediate was localized on the scaffolds. Scaffolds with 1- and 2-dimensional assemblies significantly increased pathway yields. We showed direct evidence of enzyme-RNA interaction and systematically reoriented one of the enzymes on the scaffold. We found that, in the best case, a 2D RNA scaffold led to 2.4 fold higher pentadecane production levels than observed without scaffolding. We could also demonstrate the utility of RNA scaffolds to enhance flux through 3-4 steps of a succinate production pathway with diverse intermediates.

Our system offers a convenient, modular design to optimize metabolic channeling

by changing one design element at a time. The enzymes AAR and ADO are associated with aptamers emerging on the same side of the unit tile plane, where they are well positioned to interact with each other. At each corner junction between unit tiles, 6-bp sticky ends bring together two different stem-loop aptamer segments, followed by RNA tails corresponding to the T7 terminator sequence at the 3' end of the transcripts (App. Fig A.4). Because six bases correspond to half of a turn of an RNA helix, these are likely positioned on the same side of the plane of the scaffold, and electrostatic repulsion between the aptamers, the T7 terminator RNA, and the scaffold itself will roughly determine the placement of the two interacting aptamers. In the solved structure of the PP7 protein-RNA complex (PDB 2QUX[113]) the two PP7-ADO units are at an angle of 120°. The BIV-Tat-aptamer interaction is a one-to-one interaction (PDB 1MNB[100]). Our data suggest that for some conformations of the BIV-Tat-AAR moiety that are separated by about 120°, the AAR interacts favorably with the one or the other of the PP7-ADO dimers, while other rotational conformations of the aptamer make transfer of the metabolic intermediate less favorable (Fig 3.8).

The geometric dependence of RNA scaffold performance that we report here points to a mechanism where phenomena other than just 3D spatial diffusion of metabolites[33] are in play. Our results are consistent with direct, surface-limited metabolite transfer between aligned enzyme active sites[28]. The range of flux enhancement on 2D RNA scaffolds reported here illustrates the importance of getting the precise relative orientation right when co-localizing enzymes. In the future, mathematical modeling of protein surface diffusion phenomena will shed insights into how our results and those of other groups[28,12] can be mechanistically explained. Experimentally investigating the behavior of substrates with different sizes and hydrophobicity could also help develop such a model.

Our RNA scaffolds have consistently shown better efficacy with higher dimensionalities of polymerization[133] . To understand why a similar number of aptamer-pairs result in significantly different metabolite transfer rates[13] we need a mechanism that involves transfer of intermediates not just to the nearest neighbors on a 1D or 2D scaffold, but also to downstream enzymes with more than 1 or 2 degrees of separation on the same scaffold structure. The localization of substrates to a crowded scaffold assembly would provide such a mechanism. Since scaffolds of 1 and 2 dimensions have more such neighboring pairs progressively, we would expect stronger channeling on those structures.

The economics of biofuel synthesis are particularly challenging because biofuels compete with petroleum products that derive from carbon fixed over millions of years[118]. Our results indicate that rational scaffolding of enzymes can be useful in improving yields of pathways with intermediates of a variety of different chemical properties. The 2.4-fold increase in pentadecane synthesis observed here could be synergistically combined with other genetic techniques for biofuel yield enhancement[89], leading to higher overall synthesis rates.

## 3.5 Materials and Methods

### 3.5.1 Strain construction

*E. coli* strain BL21* (DE3), that carries a mutation is the me131 (RNaseE) gene, was used for all experiments described here. The appendix Fig A.3 lists the plasmid systems used for expression of the various components in *E. coli*. RNA and protein expression was based on Novagen (EMD Millipore) Duet vectors. pETDuet1 was used for RNA expression because of its high copy number (>40 copies/cell) while the lower

copy pACYCDuet1 (10-15 copies/cell) and pCOLADuet1 (20-40 copies/cell) were used to express proteins. For co-expression, we co-transformed competent BL21* DE3 cells with the two or three relevant plasmids and selected on plates with a combination of antibiotics. For all metabolic production experiments, triplicates correspond to three independent transformant colonies.

### 3.5.2 Metabolic production culture conditions

For alkane production experiments, *E. coli* cells were grown in M9 media with 3% glucose, added minerals and 0.1% Triton X-1000 as described in Supplemental information for Schirmer et al[87]. Cultures were inoculated at OD 0.8 with 1mM IPTG. 5mL cultures were grown in 20 mL glass test tubes at 37°C with shaking. Alkane yields were measured 16 hours after induction unless otherwise stated. For succiniate production experiments, *E. coli* cells were grown anaerobically in LB media, with 2% glucose, 2% bis-tris, 100mM formate and induction with 2mM IPTG. For this, overnight LB cultures, still in exponential phase, where centrifuged and resuspended in the above media at a starting OD of 5.0 and the batch cultures were sparged to create a headspace of 16% $CO_2$ and balance Nitrogen. Succinic acid levels were measured 7 hours after anaerobic growth at 37°C.

### 3.5.3 Measuring production levels

Following induction and incubation, cell cultures were centrifuged, and the supernatant decanted. In a typical pentadecane assay, 3mL of supernatant vigorously mixed with 1mL of 1M NaCl and 1mL pure HPLC grade ethyl acetate (Sigma Aldrich). After 1 hour of shaking, the ethyl acetate extract was separated by centrifugation (10 min

at 4000g) and obtaining the top layer. Extracts were analyzed on an Agilent GC-MS 5975/7890 (Agilent Technologies) with the HP-5MS column (30 m length, 0.50 mm diameter). Following a 1-3 µl split-less injection, the inlet temperature was held at 100°C for 3 minutes and ramped up to 320°C at 30°C per minute. For quantification, standard curves of pentadecane and hexadecanol were generated, by using the pure compounds (Sigma Aldrich). Succinic acid production assays were similarly performed on decanted supernatant. Roche (r-biopharm) enzymatic succinic acid assay kits (CN: 10176281035) were used to measure production levels using a Victor3V 1420 Multilabel plate reader (Perkin Elmer).

### 3.5.4 Measuring enzyme expression

For enzyme level quantification, we collected centrifuged cell pellets from cultures that were grown in media condition described for alkane production. Cultures grown for 16 hours after induction were used. The cell pellets were lysed with a bacterial protein extraction reagent (B-PER II, Thermo Scientific), boiled for 5 minutes at 100°C, and a protease inhibitor cocktail (Roche) was immediately added. We used 4-20% Tris-Glycine gels (Novex) to separate the proteins by mass, transferred them to a nitrocellulose membrane (Novex iBlot) and blotted with antibodies. We used Abcam antibodies for endogenous GAPDH (ab85760) and 6x His Tag (ab1187), and Novagen antibody for Strep Tag II (#71591). The software ImageJ[102] was used to analyze and quantify western blot images.

### 3.5.5 Purification of RNA-bound proteins

For co-purification of RNA interacting with enzymes, we performed UV crosslinking with 150 mJ/cm² of *E. coli* cultures in 6 well plates. This was followed by gentle cell lysis using B-PER I (Thermo Scientific) containing protease inhibitors, RNaseOUT (Life Technologies) and DNase I (Sigma Aldrich). The solution was subjected to 2 freeze thaw cycles (-20°C/4°C) to further ensure cell lysis. Then the lysate was centrifuged at 12000g for 10 minutes and the supernatant was processed through the HisPur Ni-NTA magnetic beads protocol (Thermo Scientific) to purify 6xHis-tagged AAR fusion protein. IgA and IgG magnetic beads (Thermo Scientific) were used as controls to pull down random proteins. The purification was verified by Western blots and then subjected to protease K (Thermo Scientific) treatment, to release any co-purified RNA.

### 3.5.6 Measuring cellular RNA Levels

Intracellular RNA was either collected from cultures grown for 16 hours after induction or from protease digestion of the purified UV crosslinked enzymes. RNA directly from cell pellets was extracted using Trizol and a Purelink RNA Mini Kit (Ambion, Life Technologies). RNA from a protease treated purified enzyme extract was extracted through acid-phenol-chloroform extraction and overnight ethanol precipitation at -20°C. From this, cDNA was prepared using SuperScript III First-Strand cDNA synthesis system (Life Technologies). We used random hexamers to reverse transcribe the entire cellular RNA pool. 500 ng of total RNA was elongated at 25°C for 10 min, followed by 50°C for 50 minutes and termination at 85°C for 5 min. RT-PCR was performed using 7.5 µl of SYBR Green Supermix (Ambion), 4.5 µl of 10x diluted cDNA and 3µl of 2nM primer solution. Fig A.5 in the appendix lists the primers that

were used for amplification. The gapA primers served to enable measurement of endogenous cellular gapA mRNA pools, which were used as internal controls. The 2D primers shown here amplified scaffold sequences accurately as tested by a standard curve using 4 fold dilutions of a single test sample. Quantitative PCR was performed on an Eppendorf Mastercycler ep realplex thermal cycler (upto 55 cycles) and the qPCR software Biogazelle Plus was used for calculating scaffold RNA concentration relative to gapA mRNA. Quantitative PCR efficiency was estimated to be 1.87/cycle for scaffold RNA and 2.03/cycle for gapA using a dilution series. Hence, relative amplification, r, for scaffold/gapA per cycle is $r \approx 0.92$.

### 3.5.7 Scaffold structure representation

The Make-NA tool (James Stroud 2004, 2011), available on the University of Southern California servers, was used for constructing PDB representations of our stem and T7 terminator hairpins. The software Autodesk Maya (student version 2014) was used to render the 3D representation in Figs 3.7 and 3.8, along with a Molecular Maya toolkit (Digizyme).

*Cyanobacteria are some of the simplest microbes capable of harnessing energy from sunlight and are among the most evolutionarily ancient organisms. Indeed, cyanobacterial biochemistry, acting over millions of years, radically altered the atmosphere and enabled life as we know it on Earth today. It seems therefore fitting that cyanobacteria could be an important component to help us address some of today's urgent energy and atmospheric problems.*

Danny Ducat, Michigan State University

# 4

# Controlling metabolic fluxes in cyanobacteria

Cyanobacteria are ubiquitous photoautotrophic bacteria that convert solar energy into fixed electrons at 3-12 times higher efficiencies than land plants[119,91] while also being relatively easy to engineer[90]. This makes them very exciting candidates

as platforms for the renewable production of both a variety of high value chemicals[91] and feedstocks usable as substrates for bioreactors[120]. To that end, in this chapter I present research efforts aimed at understanding the control of electronic and metabolic fluxes in the cyanobacterium *Synechococcus elongatus*.

In section 4.1 I present work where we* elucidated the flow of reducing equivalents in engineered strains of *S. elongatus* and showed that electron transport reactions in cyanobacteria can be employed for renewable hydrogen production. In section 4.2, I present a design strategy we† implemented with the goal of expressing 2D RNA scaffolds (discussed in Chapters 2, and 3) in *S. elongatus* and using them to carry out the carbon fixation reaction naturally occurring within protein shells called carboxysomes.

## 4.1 Rewiring hydrogenase-dependent redox circuits in cyanobacteria

It has been shown that photosynthetic organisms are capable of hosting functional hydrogenase enzymes, carrying out the reversible reaction, $2H^+ + 2e^- \rightarrow H_2$[121]. There are two classes of hydrogenase enzymes found naturally, nickel-iron [NiFe] and iron-iron [FeFe], classified based on the nature of the reactive metallocluster at their active sites[121]. The two kinds of hydrogenases are phylogenetically and structurally unrelated. [NiFe] hydrogenates are found in a variety of organisms, however [FeFe] hydrogenases are typically prokaryotic, or found in some algal species. [NiFe]-hydrogenases are typically exchange electrons with NAD(P)H, (reducing potential $\approx$ 320 mV). [FeFe] hydrogenases interact with ferredoxins, which are electron bearing proteins and have

*The research presented in section 4.1 was carried out with lead author Daniel C. Ducat and the material has been published[86]

†Work presented in section 4.2 was carried out with Simon Alexander Thomas Kretschmer

significantly lower redox potentials, which brings the potential energy of their electrons closer to the redox potential of the $H_2/H^+$ pair ($\approx$ 420 mV)[122]. Thus [FeFe]-hydrogenases thermodynamically favor production of $H_2$ compared to [NiFe] hydrogenases, which are predominantly carry out $H_2$ uptake[123].

Hydrogenases, because of their metallocluster active sites are usually oxygen sensitive, and thus it can be a challenge to employ them for biological hydrogen production in photosynthetic organisms[124]. Several efforts to this end have used various strategies to temporally insulate the hydrogenase reactions from the water splitting (hence $O_2$ producing) activity of the photosystem II (PSII). This involves growing photosynthetic cultures normally to allow generation of internal stores of reducing equivalents, then changing the cellular environment of one that is both anaerobic and inhibits PSII activity[125]. Both, anaerobic light fermentation[126], and dark fermentation[127] have been shown to result in biological hydrogenase activity leading hydrogen production.

Here, we expressed a *Clostridial* [FeFe] hydrogenase HydA[128] in *Synechococcus elongatus* to achieve hydrogen production in the cyanobacterium. *S. elongatus* is obligately aerobic and photoautotrophic. It also does not contain any nitrogenases or nitrogen fixing capability[129]. Nitrogenases can sometimes carry hydrogenase activity, but any hydrogen production in our system can be solely attributed to the expressed HydA enzyme. We show that HydA can be functionally integrated into the electron transport machinery of the cell and is capable of hydrogen production. We also show that the channeling of electrons to and from HydA can be rationally designed to achieve cell growth based on hydrogen consumption and also conversely, the use of dark fermentation reactions, converting cellular carbohydrate pools into hydrogen gas.

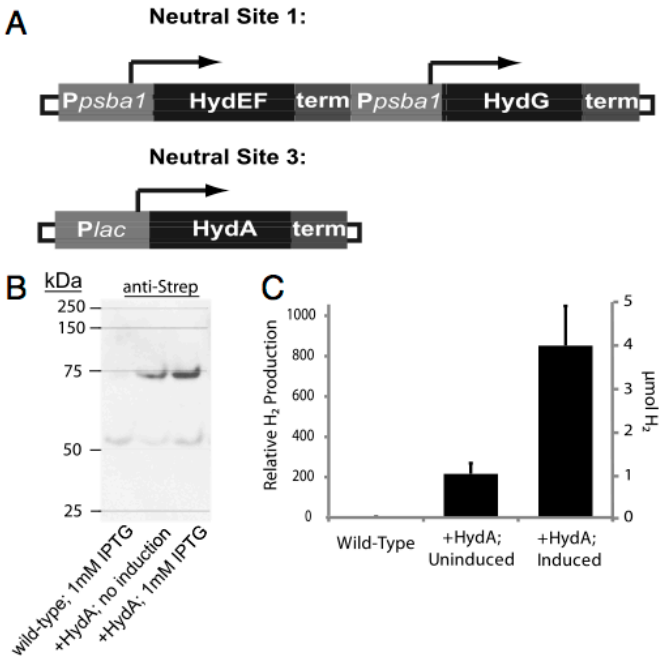



**Figure 4.1:** Expression of Hydrogenase A in *Synechococcus elongatus* sp. 7942. (A) A schematic showing the genes integrated into the genome. Constitutive promoter *psba1* are used to express maturation factors (HydEF,G) and HydA is placed under IPTG-inducible P*Lac* control (B) A western blot showing IPTG induction of HydA in transformed strains (expected molecular mass : 65.6 kDa) (C) Evolution of hydrogen gas from WT and HydA expressing strains of *Synechococcus* lysates after 60 minutes anaerobic methyl viologen treatment

### 4.1.1 Expression and *in vitro* activity of HydA

We expressed the [FeFe] hydrogenase (HydA) from *Clostridium acetobutylicum* in *Synechococcus elongatus* sp. 7942. In previous work at our lab with synthetic hydrogen-producing circuits in *Escherichia coli*, the HydA from *C. acetobutylicum* was shown to have the highest level of activity among several hydrogenases tested[130].

We also expressed hydrogenase maturation factors HydEF and HydG[131], by cloning them downstream of the *S. elongatus* psba1 promoter. The psba1 promoter controls the expression of D1 subunit of photosystem II (PSII), leading to constitutive expression. This cloned unit was integrated into the cyanobacterial genome using the neutral site

1[90] which is a genomic location that has been shown to allow genetic integration without any noticeable phenotypic effects (Fig 4.1A). HydA was similarly inserted at the neutral site 3[132] (Fig 4.1A). HydA was however placed until the IPTG inducible PLac promoter. Expression of a StrepII-tagged HydA could be confirmed by Western blot analysis and the protein expression could be induced (approximately 3-4 fold) by addition of IPTG (Fig 4.1B). Next, we used a standard methyl viologen assay[128] to confirm the catalytic activity of the expressed enzyme.

Methyl viologen is a promiscuous electron donor that can reduce both [FeFe] and [NiFe] hydrogenases. Thus, if cell lysates are mixed with an excess of methyl viologen, the hydrogen gas evolved is an accurate measure of the hydrogenase activity[128]. We incubated cell lysates with methyl viologen in an anaerobic environment for 45 minutes and observed >500-fold increase in the amount of hydrogen gas detected in the headspace (Fig 4.1C) with HydA expressing cultures. Under similar conditions, wild-type (WT) cyanobacterial cultures yielded no $H_2$ gas.

### 4.1.2 *In vivo*, light dependent hydrogen evolution

To achieve functional hydrogenase activity *in vivo*, the enzymes must interact with the endogenous redox machinery and in order to be able to accept reducing equivalents. However, due to the extreme oxygen sensitivity of HydA, and most hydrogenases in general, it is necessary to spatiotemporally separate the water- splitting reaction of PSII from hydrogenase expression. We used 5 µM diurion [3-(3,4-dichlorophenyl)-1,1-dimethy- lurea (DCMU)], to inhibit PSII activity after cells had been grown to a healthy density (OD750 ≈ 0.5-1). At this point we also sparged the cyanobacterial cultures' headspace with 2.5% CO2 (N2 balance) to create an anaerobic environment. As before, expression of hydrogenase was induced with IPTG. We then incubated the cul-

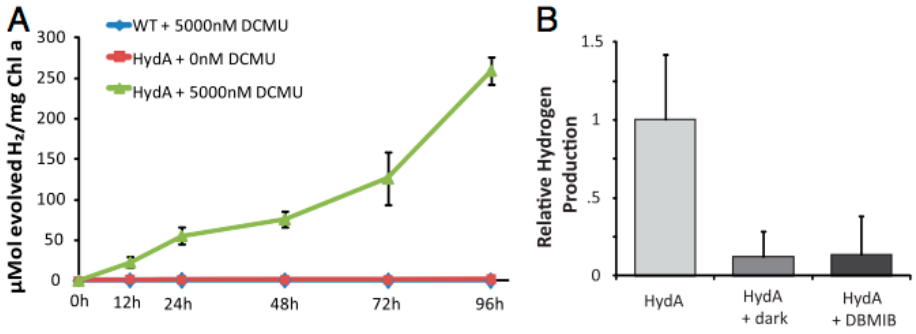



**Figure 4.2:** *In vivo* hydrogen production from HydA expressing *Synechococcus elongatus*. (A) Hydrogen production under anaerobic conditions with dichlorophenyl dimethyl urea (DCMU) added to media to prevent oxygen evolution from photosystem II (B) *In vivo* hydrogen production depends on light and electron transport from plastoquinone, which can be inhibited by dibromothymoquinone (DBMIB).

tures with or without light and assayed hydrogen evolution into the headspace. No hydrogenase activity was detected in the parent *S. elongatus* strain, while significant levels of hydrogen gas were produced by HydA-expressing cell cultures. Production was only observed when $O_2$ generation was suppressed (Fig 4.2A). The average measured hydrogenase activity of HydA-expressing strains over the first 96 hours was 2.8 µmol $H_2$ $h^{-1}$ · mg Chl-$a^{-1}$ (Fig 4.2A), where Chl-a is cholorphyll a, measured by methanol extraction[133].

Cyanobacteria and alga can anaerobically metabolize carbohydrates to generate reducing equivalents that may be used by hydrogenases through either light-dependent or dark reactions[134,135]. In light-dependant reactions, the electrons from carbohydrate metabolism are transferred to the plastoquinone pool, and reenergized through Photosystem I (PSI) before interacting with hydrogenases through petF- class plant-type ferredoxins[134]. During dark fermentation however, reducing equivalents from the breakdown of glycogen can either be directly donated to the hydrogenase, or may be transferred through electron carriers like NAD(P)H or ferredoxin[136,137]. We observe

that HydA-expressing *Synechococcus*, hydrogen production is greatly reduced in the absence of light or in the presence of the plastoquinone inhibitor dibromothymoquinone (DBMIB) (Fig 4.2B). Therefore, we can conclude that the electron transfer to HydA is largely dependent cyanobacterial electron transport chain function.

### 4.1.3 HydA activity supports limited growth with $H_2$ as sole electron source

In several algal species, native hydrogenases function by recovering electrons from hydrogen, that may have been produced as a side product from the promiscuity of nitrogen fixation reactions. For efficient uptake of reducing equivalents, the hydrogenases need to interact with cellular machinery in the opposite direction from the previous sub-section. They need to deposit the hydrogen-derived electrons towards the rest of cellular metabolism. It has been shown that a native bidirectional [NiFe] hydrogenase of *S. elongatus* can uptake $H_2$ and reduce $NAD^+$. This activity has been examined in cell extracts and was measured to be very low (85 nmol $h^{-1}$ · mg $protein^{-1}$)[138]. Such a rate is likely insufficient to support autochemolithic growth using $H_2$ as the only source of reducing equivalents, and $H_2$-dependent cell division has not been observed in cyanobacteria[138,139] (Fig 4.3A, WT curves).

Hence, we evaluated the the *in vivo* activity of HydA with regard to its capacity to consume hydrogen gas as an alternative source of reducing equivalents within *S. elongatus*. We tested this under conditions where PSII water-splitting reactions were inhibited by addition of DCMU. Additionally, cells were incubated in anaerobic sealed cultures with 5% $CO_2$ and 5% $H_2$ (balance $N_2$), and growth was assayed by OD750 (Fig 4.3A). We notice that HydA-expressing strains can support limited cell growth using hydrogen under DCMU inhibition when incubated anaerobically under illumination.

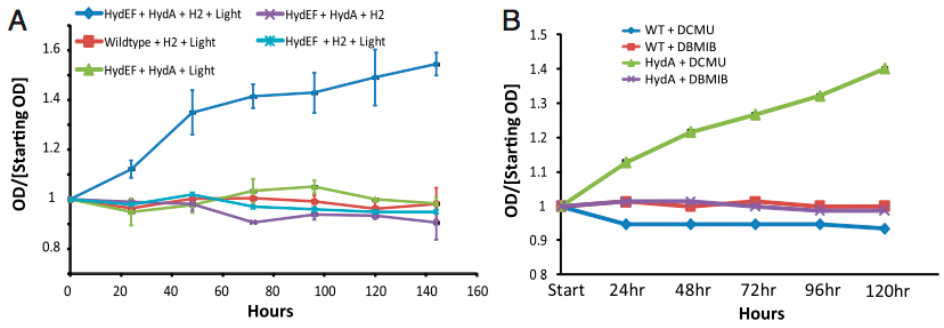



**Figure 4.3:** Hydrogenase A dependent chemoautotrophic growth of *Synechococcus*. (A) HydA activity supports growth in anaerobic conditions with hydrogen gas and co-expression of maturation factors HydEF (B) Using dibromothymoquinone (DBMIB) to inhibit plastoquinone electron transfer prevents hydrogenase mediated growth

Most of the growth was observed in the first 3-4 days following the switch to a hydrogen containing atmosphere, and HydA-supported growth depended on the presence of $H_2$, the absence of $O_2$, HydA maturation factors, and illumination (Fig 4.3A).

Carbon fixation through the Calvin cycle requires supplies of both ATP and reducing equivalents in order to increase cell mass. Under illuminated conditions, photosynthetic cells carry an excess of reducing equivalents relative to ATP, and electrons from ferredoxin can reenter the ETC through the plastoquinone pool in an essential process termed cyclic electron transport[140]. Although cyclic electron transport increases the proton gradient across thylakoid membranes, hence, stimulating ATP production, light absorption at PSI is essential to keep ferredoxin pools reduced, for supplying electrons to Calvin cycle. Oxidation of hydrogen, however, can only provide a source of electrons. Based on the observation that HydA-dependent growth required light as well, we hypothesized that hydrogenase-derived electrons were being channeled through the ETC and undergoing cyclic electron transport to create proton gradients and generate ATP. To test this, we attempted to block electron flow, using the plastoquinone inhibitor DBMIB. We noticed that HydA-dependent growth was disrupted

when electron flow from plastoquinone was blocked by the addition of DBMIB (Fig 4.3B). We also observed that HydA-expressing strains were more viable after extended incubations under the above conditions (DCMU/hydrogen), as evident by culture bleaching and colony forming unit (CFU) counts when plated on agar plates. Taken together, these results show that heterologously expressed *Clostridial* HydA is capable of both accepting and donating reducing equivalents from/to the endogenous redox circuitry of *S. elongatus*.

### 4.1.4 HydA activity depends on the co-expressed ferredoxins

As mentioned earlier, it is likely that HydA is interacting with endogenous *S. elongatus* petF plant-type ferredoxins, leading to $H_2$ production that is dependent on both light and ETC function (Fig 4.2B). However, it is also possible that the endogenous plant-type ferredoxins may be only weakly interacting with the heterologous HydA. In a previous project in the lab, employing a synthetic circuit in *E. coli* to test interactions of HydA with a several ferredoxin proteins, it was found that some ferredoxin-hydrogenase affinity pairs could influence levels of hydrogen output from the circuit[130]. We tested these pairs to facilitate enhanced electron transfer to HydA and increase hydrogen production from *S. elongatus*. The strongest ferredoxin pairs, one plant type and one bacterial type (from *Spinacia oleracea* and *C. acetobutylicum* , respectively)[130], were integrated into the genome of the HydA-expressing strain, and were also cloned downstream of an IPTG-inducible promoter[90]. We found that co-expression of the *C. acetobutylicum* (CAC0303) ferredoxin[141], the strongest pairing from previously reported synthetic circuits[130], increased the rate of H2 production by nearly twofold (Fig 4.4A). This suggests that expression of the *Clostridial* ferredoxin helped to redirect cellular reducing equivalents toward HydA more efficiently. We found that co-expression

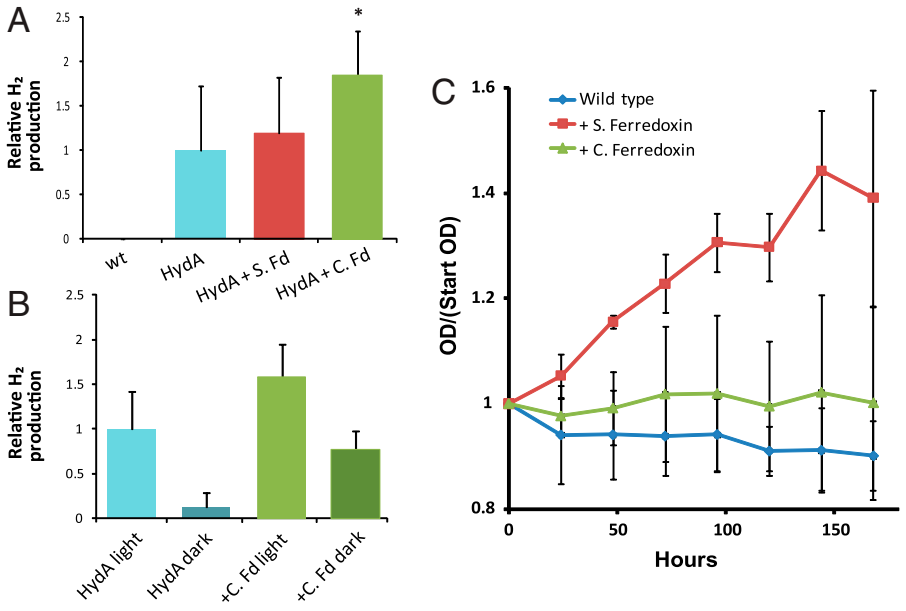



**Figure 4.4:** Coupling with ferredoxin affects HydA electron flow. (A) Effect of co-expressing ferredoxins from *Spinacea oleracea* (S. Fd.) and *Clostridium acetobutylicum* on hydrogen production through HydA (B) C. Fd. enables hydrogen gas production from dark fermentation (C) Strains expressing C. Fd. are unable to support HydA dependent growth in an anaerobic hydrogen environment.

of the plant-type ferredoxin I from *S. oleracea* did not significantly enhance hydrogen production.

Although the ferredoxin from *C. acetobutylicum* has previously shown a strong functional interaction with HydA in a synthetic circuit[130], it is a bacterial-type ferredoxin and is unlikely to be capable of direct interactions with PSI in a manner similar to the endogenous *S. elongatus* plant-type ferredoxins. Therefore, we asked if the observed increase in hydrogen production in strains expressing both HydA and *C. acetobutylicum* ferredoxin was the result of utilization of reducing equivalents from a distinct source. Indeed, *C. acetobutylicum* ferredoxin-bearing strains were capable of producing hydrogen in the dark and/or in the presence of DBMIB, whereas these conditions largely abolished hydrogen production when expressing HydA alone (Fig

4.4B).

Since the heterlogous expression of a ferredoxin enhanced the ability of HydA to produce hydrogen (Fig 4.4A), we examined if new ferredoxin pairings would affect the ability of HydA to support hydrogen dependent anaerobic growth as well (Fig 4.3A). We found that strains co-expressing *S. oleracea* ferredoxin showed HydA-supported growth in the presence of hydrogen as before (Fig 4.4C). However, remarkably, the presence of *C. acetobutylicum* ferredoxin attenuated $H_2$-dependent growth (Fig 4.4C).

Together, these experiments imply that co-expression of exogenous ferredoxins can both increase the flux of electrons toward HydA, and "rewire" the redox pathway, depending on the specific ferrodoxin-HydA links being optimized. Our hypothesis is that the *C. acetobutylicum* ferredoxin provides an alternative electron circuit that bypasses the need for light reactions and anaerobic fermentation through PSI for the reduction of HydA. However, since this exogenous ferredoxin, which is presumably a preferred interacting partner of HydA, appears less integrated with the cyanobacterial redox pathways, it may "short circuit" the HydA interactions with native machinery that were responsible for $H_2$-supported growth (Fig 4.5).

### 4.1.5 HydA expression likely creates new electronic flows in *Synechococcus elongatus*

In this work[86], we demonstrated the functional expression of an exogenous [FeFe] hydrogenase (HydA) in cyanobacteria *in vivo*. In particular, we showed activity within the context of the endogenous redox machinery and the electron transport chain. We show significantly enhanced hydrogen evolution rates relative to those of wild-type *S. elongatus*, which rely on native [NiFe] hydrogenases (Figs 4.1C and 4.2A). It is relevant to note that in an earlier study attempting to use a [FeFe] hydrogenase from *Clostrid-*

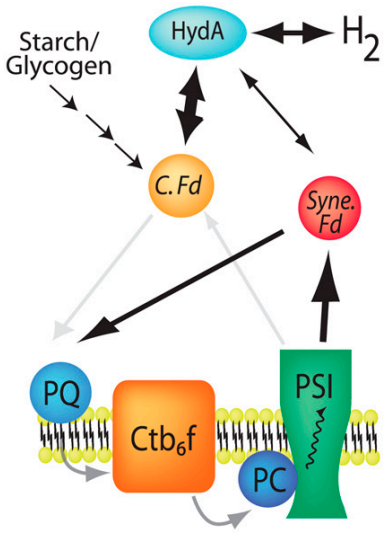



**Figure 4.5:** Model of ferredoxin-mediated HydA electron flows in *Synechococcus elongatus* We hypothesize that cyanobacterial endogenous ferredoxins (Syne. Fd) mediate electron transfer from photosystem I to hydA, enabling both light dependent hydrogen production, and hydrogen dependent growth. C. Fd can interact more strongly with HydA and also likely picks up electrons from dark reactions. However, C. Fd interactions with plastoquinone are inefficient (faded arrows), preventing hydrogenase dependent growth.

*ium pasteurianum* within cyanobacteria[142], the hydrogenase exhibited only limited (nearly two-fold) increase in activity over the native hydrogenases. Here, we show that an HydA is capable of interacting with endogenous redox circuits to evolve hydrogen from cellular sources of reducing equivalents at 2.8 μmol $H_2$ $h^{-1}$ · mg Chl-$a^{-1}$ (Fig 4.2A). We also show that HydA can support chemoautotrophic growth of *S. elongatus* by uptake of hydrogen gas as the sole source of reducing equivalents when photosystem II is chemically inhibited.

Our lab has previously shown that *C. acetobutylicum* HydA can functionally interact with a variety of ferredoxin proteins from diverse organisms[130], and here we show that electron carriers within *S. elongatus* can also donate and/or receive reducing equivalents to/from HydA. Additionally, our results show that selected ferredoxin-

hydrogenase combinations may optimize hydrogen production and change input/output sources for the reaction, such as allowing for fermentation of internal reducing equivalents without the assistance of light (Figs 4.4B and 4.5). Although the presence of distinct (i.e., light-dependent and light-independent) pathways for electron transfer to hydrogenases has been documented[137,143], to our knowledge this is the first demonstration of genetic means to enhance the rate of one pathway relative to the other. Rewiring the flow of electrons by such a strategy may open up the possibility of implementing parallel redox pathways that are insulated from host redox machinery (Fig 4.5). Our results suggest that characterization and optimization of the key residues forming the ferredoxin-hydrogenase interface could be a big step towards deterministic control over novel electron pathway design.

Employing cyanobacteria for the biological synthesis of hydrogen gas and other high value chemicals is an environmentally friendly approach to meet future energy and material needs[91]. In this work, strains expressing [FeFe] hydrogenase showed a >500-fold increase in hydrogen production *in vivo* relative to unmodified strains of *S. elongatus*[139]. A successful hydrogen producing organism would likely need to be both photoautrophic and need minimal additional care in the form of bioreactors. In this regard, discovery or engineering of hydrogenases with increased oxygen tolerance would enable direct synthesis of hydrogen gas, bypassing inefficiencies associated with energy transfer and storage through carbohydrate intermediates. The ability of HydA to oxidize hydrogen to support limited growth in *S. elongatus* also offers a possible method for directed evolution of HydA variants with desirable traits, particularly greater oxygen tolerance. To our knowledge, there have been no previous reports of cyanobacteria capable of chemoautrophic division through the utilization of hydrogen. Given the relative ease of genetically modifying *S. elongatus*, it may be possible to make further

improvements to this strain similar to those shown in other cyanobacteria or algae that enhance starch accumulation[144], reduce cyclic electron transport[126], modify PSII activity[145], or reduce phycobilisome size[146] to allow for increased hydrogen output. Finally, optimal culture conditions and cell immobilization may further enhance the yield of photobiologically produced hydrogen[147,148].

### 4.1.6 Materials and methods

#### Strains, Plasmids, and Culture Conditions

Wild-type *S. elongatus* was obtained from the American Type Culture Collection. Cultures and in vivo assays utilized BG11 medium with light illumination of 4000 lux at 30°C and a 12/12 h day/night cycle. All constructs were cloned in BioBrick format in *E. coli* (Assembly Standard 21)[149]. *S. elongatus* cultures in log phase (OD750 ≈ 0.4) were transformed with 100 ng plasmid DNA overnight and plated on BG11 plates with antibiotics. Neutral site vectors for genomic integration were obtained from their respective lab of origins [NS1/pAM2314; NS2/pAM1579[90,150]; NS3/ pHN1-LacUV5[132]]. Hydrogenase (HydA) from *Clostricium acetobutylicum*, *Clostridium acetobutylicum* ferredoxin CAC0303, ferredoxin I from *Spinacia oleracea*, and hydrogenase maturation factors (HydEF, HydG) from *Chlamydomonas reinhardtii* were cloned and synthesized as previously described[130].

#### Hydrogen production and uptake assays

*S. elongatus* was grown under a standard atmosphere with appropriate antibiotics, diluted to OD750 ≈ 0.1 at volumes 10-50 mL within 25-240 mL clear glass vials and sealed with rubber septa. Where applicable, the following chemicals were supplemented in

the media: 5 μM Diuron (DCMU), 20 μM DBMIB, 0.1-1.0 mM IPTG, 20 mM HEPES pH 8.0. Cultures were sparged with 2.5% $CO_2$, nitrogen balance (Airgas) and placed in 30°C incubators with light as above (or dark, where appropriate). For in vitro methyl viologen assays, we adapted the method from King et al[128], lysing cyanobacterial sealed cultures with bacterial protein extraction reagents (Pierce) in the presence of 1 mM methyl viologen (Sigma) and 5mM sodium dithionite for 60 min (Fisher). Headspace was measured by gas chromato- graphy (Shimadzu GC-14A). For hydrogen-dependent growth assays, headspace was sparged with 5% $H_2$ , 5% $CO_2$ , nitrogen balance.

## 4.2 Towards a synthetic $CO_2$ fixation module based on RNA scaffolds

Ribulose- 1,5-bisphosphate carboxylase/oxygenase (rubisco) is the most abundant protein on the planet. It catalyses a key step in the fixation of carbon dioxide into carbohydrates as a part of the Calvin cycle, hence feeding nearly all life on earth. It is widely held that Rubisco's slow catalytic rate of 3-10 molecules of $CO_2$ per second, which likely explains its abundance, "is a relic of a bygone age"[151] and that the 'inefficiencies' of the enzyme are due the primitive environmental gas compositions at the time of its origin. One of the key inefficiencies, is a competing oxygenation reaction where $O_2$ can react with the 5 carbon substrate at rubisco's active site, causing losses of upto 25% of fixed carbon in some species[152] through a process called photorespiration. Photosynthetic organisms have evolved several adaptations to counter this loss by protecting Rubisco through carbon concentrating mechanisms like C4 photosynthesis[152] in arid land plants and carboxysomes in cyanobacteria[153].

Carboxysomes are protein micro compartments in cyanobacteria[21] that co-localize

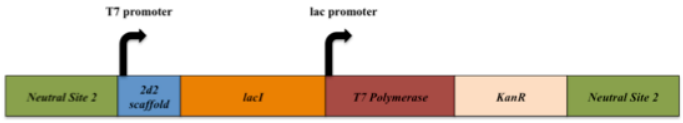



**Figure 4.6:** Schematic for inserting 2D RNA scaffold expression module into *Synechococcus elongatus*. This module includes the T7 RNA polymerase needed for RNA scaffold strand expression and is designed to be targeted to the neutral site 2 region [90] of *S. elongatus*

carbonic anhydrase and rubisco [154]. Carbonic anhydrase is a part of the carbon concentrating mechanism in cyanobacteria and catalyses the conversion of dissolved (and concentrated) bicarbonate into gaseous $CO_2$, thus providing a high local concentration of $CO_2$ relative to oxygen inside the carboxysomes. Carboxysomes are complex compartments - their biogenesis is only starting to be elucidated [155,156], and attempts have been made to express them in other hosts, like *E. coli* [157].

There is some debate about the necessity [158] and mechanisms [25] of selective permeability of the carboxysomes for $HCO_3^-$, $CO_2$, and $O_2$. We decided to examine whether the carbon concentrating effect of carboxysomes can be mimicked on the *in vivo* RNA scaffolds [13] described in Chapters 2 and 3. For this, we expressed the RNA scaffolds in cyanobacteria alongside with fusion proteins of rubisco and carbonic anhydrase with aptamer binding domains - to co-localize the enzymes, and then tested the function of a proposed assembly in strain where natural carboxysomes are not functional.

### 4.2.1 Expressing RNA scaffolds in *S. elongatus*

The first step was to achieve expression of RNA scaffold strands in *S. elongatus*. The Fig 4.6 shows the gene construct used to achieve this. In the figure, the green regions labeled *Neutral Site 2* were targeting regions to achieve genomic integration of the construct [90]. The scaffold expression system consists the 2D RNA scaffold strands with

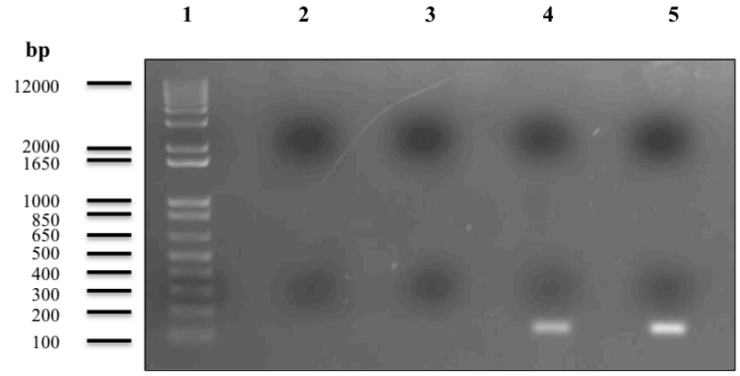



**Figure 4.7:** Expression of scaffold RNA in *Synechococcus elongatus*. The different columns are (1) 1kb+ DNA ladder, (2) WT *S. elongatus* with no IPTG, (3) WT *S. elongatus* with 1 mM IPTG, (4) *S. elongatus* transformed with 2D scaffold - no IPTG, and (5) *S. elongatus* transformed with 2D scaffold - 1 mM IPTG.

MS2 and PP7 aptamers[13] under the control of a PLac-T7 promoter[85]. The T7 RNA polymerase - promoter system has a short GGG pre-mRNA transcription start sequence[159] that allows for better control over the expressed scaffold sequence folding. To allow use of this, we included the T7 polymerase under the control of an IPTG inducible PLac system. Since both promoter cloned in were controlled by the lac operator, we included a constitutive LacI expression sequence into this construct. Finally, Kanamycin antibiotic resistance was included to be able to select for the transformed colonies.

We tested expression of the RNA scaffolds after integration of the construct into *S. elongatus* by reverse transcriptase PCR and agarose gel analysis of the cDNA. We used random hexamers for cDNA preparation and the primers (Fig A.5) for 2D scaffold, also used in Chapter 3, for amplification of the cDNA of interest (Fig 4.7). We can note from the columns 4 and 5 of Fig 4.7, that the RNA scaffold strands could be inducible expressed in *S. elongatus* using the designed expression system (Fig 4.6).

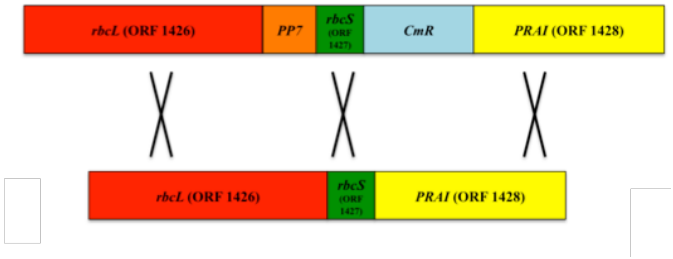



**Figure 4.8:** Schematic for inserting a PP7-rbcS (rubisco small subunit) fusion in *Synechococcus elongatus* sp. 7942 at the rbcS native gene operon location

### 4.2.2 Inserting RNA binding domain fusions into the genome of *S. elongatus*

We decided to make fusions of rubsico with PP7 and carbonic anhydrase with MS2, for co-expression and localization with the 2D scaffold. In order to mimic natural function, we decided to express these fusions by inserting the constructs into the genomic locations for rubisco and carbonic anhydrase respectively, with minimal perturbation of the genomic operons barring the fusion protein insertions and selection markers.

#### Rubisco fusion with PP7

We designed and cloned to construct shown in Fig 4.8 to express a fusion of rubisco with the PP7 domain in order to localize it to the 2D RNA scaffold. The functional rubisco holoenzyme is a hexadecamer composed of eight large (rbcL) and eight small (rbcS) subunits[160,161]. We studied the rubsico holoenzyme crystal structure (PDB:1GK8[161]) to decide on making a N-terminal fusion of the rbcS subunit with PP7. This decision was supported by a report where an rbcS N-terminal fusion with GFP was shown to assemble and retain catalytic function as a part of the rubisco holoenzyme and within active carboxysomes[162]. To enable the homologous recombination mediated genomic

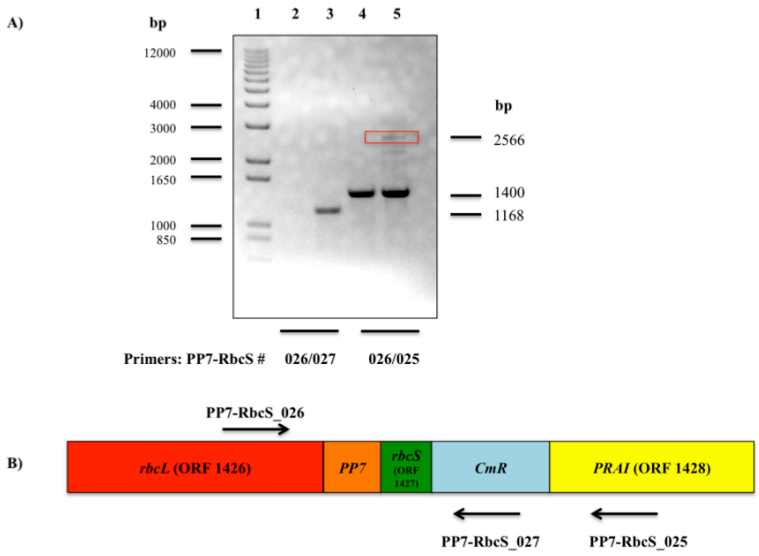



**Figure 4.9:** Verifying the insertion of Rubisco-RNA binding domain fusion in *Synechococcus elongatus* sp. 7942. This gel shows products from a genomic PCR of the WT (columns 2 and 4) and transformed chloramphenicol resistant (columns 3 and 5) *S. elongatus* strains. Primers 026 and 027 are only expected to yield a band from the inserted fusion constructed that includes a chloramphenicol resistance marker. Primers 026 and 028, however would yield a product of length $\approx$ 1500 b.p. from the WT genome, and one of length 2566 b.p. from the transformed strain.

integration of this construct in a way that replaces endogenous gene with the fusion (genomic knock in - Fig 4.8), we included the upstream and downstream genomic sequences, rbcL and PRAI respectively. Chloramphenicol resistance was used as a selection marker.

In cyanobacteria, often most of the population consists of cells with 3-6 copies of the genome[3] . It is common to verify that colonies produced after a transformation are completic genonmic knock-ins, i.e. all the genomic copies have been altered. When we carried out such a test on the rbcS-PP7 fusion insertion into the genome (Fig 4.9) we found the despite the fact the colonies and cultures were all chloramphenicol resistance, only a small fraction of the genomes took up our PP7 fusion insertion. In Fig 4.9, this can be seen from the PCR products on the gel. In column 5, the PP7 fusion


**Figure 4.10:** Schematic for inserting chloramphenicol resistance marker gene into *Synechococcus elongatus* to generate a strain which could control for growth rate changes due to presence of an antibiotic and a marker gene.

would have resulted in a product of length 2566 bp using the primer locations shown, while the WT genomic copy would yield a band of length nearly 1500 bp. We see that while the PP7 insertion is present in the transformed cells (Fig 4.9A - column 3), the overwhelming majority if the locations are still WT (Fig 4.9A - column 3). We hypothesized that this reflects a severe fitness disadvantage on creating a rbc-PP7 fusion, which might not function at optimal rates in the carboxysome. We maintained this stain with chloramphenicol, and wanted to test if including a 2D scaffold and localization with carbonic anhydrase (CA) might help rescue the fitness for PP7-rbcS fusions.

However, we also wanted to test whether the current strain (a mix of WT and PP7-rbcS genomes) were already showing a fitness disadvantage. To test this, we cloned a control construct, which expresses only the chloramphenicol resistance gene at the neutral site 3 (Fig 4.10) and compared the its transformants with those tested in Fig 4.9. The growth curves in Fig 4.11 show that no growth rate difference was noticeable and that the current mixed strain maintained close to WT growth rates.

### Carbonic anhydrase fusion with MS2

The genome of *S. elongatus* PCC 7942 carries three ORFs labeled with putative carbonic anhydrase activity. ORF 1388 codes for ecaA, an $\alpha$-CA that localizes to the periplasm and is believed to play a role in the regulation of $CO_2/HCO_3^-$ as a part of the carbon concentrating mechanism[163]. On the other hand, ORF 1447 is the thought to code for the major carbonic anhydrase ccaA associated with cyanobacterial carboxysomes[154].

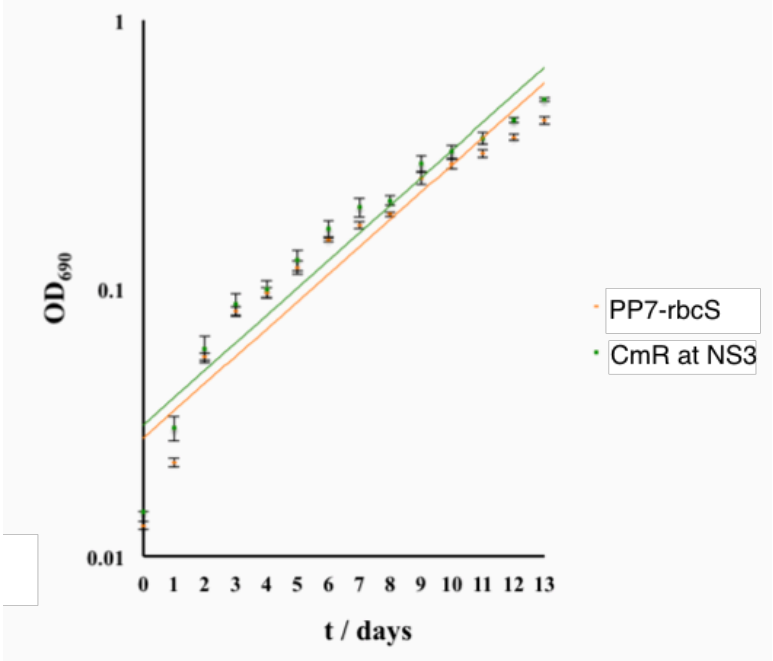



**Figure 4.11:** Comparing growth rates for transformed strains of *Synechococcus elongatus* - the green fit shows estimated log phase growth for a strain carrying just the chloramphenicol resistance marker, and the orange fit is growth of the mixed PP7-rbcS/WT genome carrying chloramphenicol resistant strain. The slopes of the two are the same - revealing comparable growth rates.


**Figure 4.12:** Schematic for inserting ccaA-MS2 fusion into *Synechococcus elongatus* at the ORF 1447, i.e. the native genomic location for the ccaA gene.

Finally, the protein ccmM, known to be a carboxysomal shell component, also bears homology to archaeon ɣ-CA at its N-terminus, though no direct evidence of its catalytic activity exists[164].

Thus, we picked the protein ccaA (encoded by ORF 1447) to make a C-terminal fusion with the MS2 aptamer binding domain (Fig 4.12). As before, the upstream and downstream sequences (ORFs 1447 and 1448) were flanking the MS2 fusion region and a spectinomycin resistance marker, to allow genomic integrations (Fig 4.12). Unlike with the rubisco fusion, this time we were able to achieve complete integration of the fusion into all copies of the genome. Once again, the strains carrying these fusions of ccaA-MS2 did not show any growth defect compared to WT.

### 4.2.3 Testing carbon fixation on scaffolds in a carboxysome-affected strain

Next, we needed to create a strain with dysfunctional carboxysomes, so that we could show that localizing of rubisco and carbonic anhydrase on 2D RNA scaffolds could rescue their growth phenotype. The dysfunction of central components of the carbon concentrating mechanism and the carboxysomal shell proteins results in a high carbon

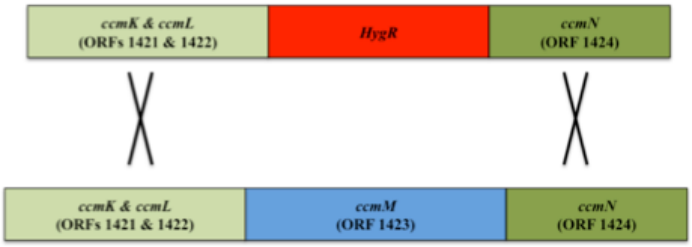



**Figure 4.13:** Scheme to knock out carboxysomal shell protein ccmM from *Synechococcus elongatus* by inserting a selectable hygromycin resistance marker in its genomic location

requiring (hcr) phenotype. The hcr phenotype is characterized by the need for elevated levels of $CO_2$ in media (or environment) to support growth[160]. It has been shown that specifically, the deletion of a protein discussed above, ccmM leads to a the high carbon requiring phenotype since ccmM is essential for functional carboxysomal assembly[165].

We designed the construct shown in Fig 4.13 to knock out the ccmM expression (ORF 1423) and replace it with a selectable hygromycin antibiotic resistance marker. We could show that this strain had a high carbon requiring phenotype, as seen in Fig 4.14. As can been seen in Fig 4.14A that the growth rate for the hygromycin resistant strain with ccmM knocked out has a growth rate in air that is significantly lower than the WT growth rate. However, the strain's growth rate is comparable to WT levels under high carbon (2% $CO_2$) conditions (Fig 4.14B).

With all the parts in place, we decided to test whether co-expression of the PP7-rbcS and ccaA-MS2 fusions with the 2D scaffold strands carrying PP7 and MS2 aptamers will be able to rescue the high carbon requiring phenotype of a ccmM knockout strain. For this, a strain was created that contained all of the mutations described above and was tested for growth rate, with controls, both in air and high $CO_2$ conditions (Fig 4.15). As expected, all strains showed healthy growth rates in 2% carbon dioxide envi-

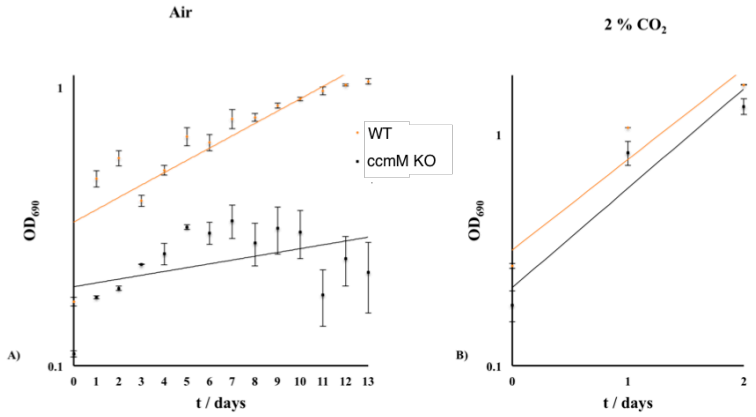



**Figure 4.14:** High carbon dioxide dependent growth in *Synechococcus elongatus* lacking carboxysomal shell protein. The orange curves are fits for WT growth and the black curves fit growth of the ccmM knock out strain. (A) Growth measured in batch culture in flasks grown in air, (B) growth measured from batch culture flasks in a 2% $CO_2$ incubator

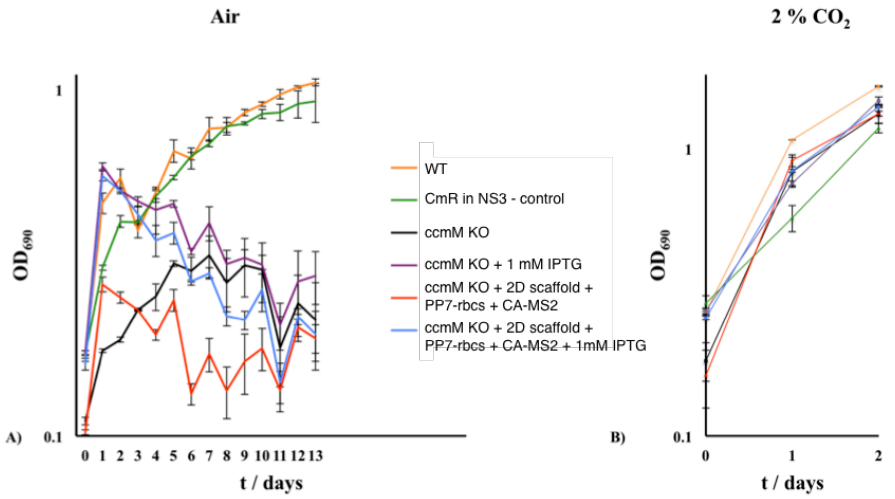



**Figure 4.15:** Testing localization of carbon fixation on RNA scaffolds in *Synechococcus elongatus* lacking carboxysomal shell protein. Growth rates for the various positive control (WT, and cmR), negative control (ccmM KO) and tested (ccmM KO + 2D scaffold + PP7-rbcS + ccaA-MS2) strains are shown from batch cultures in (A) air, an (B) 2% $CO_2$

ronment (Fig 4.15B). WT and the chloramphenicol resistant strains showed normal growth rates in air (Fig 4.15A). The ccmM knock out strain, expected to be the negative control here, did not grow in air. However, the ccmM knock out strain carrying PP7-rbcS, ccaA-MS2 fusions, and with the scaffolding system induced with 1mM IPTG, did not show growth either (Fig 4.15A). Hence, in it's first pass, the system was unable to mimic carboxysomal function and rescue the hcr phenotype of a ccmM knockout strain. We believe that even there can be several ways to design this system more effectively and there are also likely to be some conditions that are less stringent than air, where such a system may show an effect. Regarding the former, since the rubisco hexadecamer is sterically larger than most holoenzymes (at nearly 8 nm across), is likely that the current 2D scaffold design, with just a 6bp-linker (see Fig 3.7) connecting the two aptamers, might be too crowded to see effective distribution of rubicso with carbonic anhydrase. Even if the 2D scaffold is able to channel some of the $CO_2$ locally to rubisco holoenzymes associated with the 2D scaffold, the increase in local concentration may not be high enough for rubisco to preferentially interact with $CO_2$ instead of $O_2$. Testing a new set of designs at intermediate $CO_2$ concentrations (between 0.04% and 2%) may be important to find a regime where function is revealed.

### 4.2.4 Directions

In this section we showed that RNA scaffolds can be expressed in *S. elongatus* and attempted to use them for demonstration of carbon dioxide channeling. The T7 RNA polymerase based RNA scaffold system can be generally used for many other applications in cyanobacteria. RNA based tools for sub-cellular organization can help us probe unanswered questions in basic biology[166], customize metabolism in synthetic compartments[167], or even help us be able transfer efficient bacterial mechanisms like

carbon concentrating, into plants to increase their carbon fixation efficiencies[168]. As we have seen in Chapter 3, RNA scaffolds offer customizable control over precise location of, and flux through co-localized enzymes, which when combined with a photosynthetic chassis[120] opens doors for cheap, renewable, production of biochemical products directly in oceanic or marine environments.

*There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one: and that, whilst this planet has gone circling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.*

Charles Darwin, *On the origin of species by means of natural selection*[169]

# 5

# Conclusion

Today, synthetic biology has brought us to the brink of creating life itself[170]. Evolution through natural selection[169] provides us with many diverse templates revealing how biological systems achieve complexity[171], robustness[172], and adaptive increases in efficiency[152]. Here we have demonstrated means of engineering bacterial systems with biologically inspired design to achieve increased metabolic outputs.

We were able to expand the toolkit for expressing *in vivo* RNA scaffolds for pro-

teins. Complementation of split GFP fragments served as a reliable assay for characterizing function of these new scaffolds in bacteria. We then went on to localize metabolic enzymes from two different enzymatic pathways inside cells. Our scaffolds were able to tether 2-4 enzymes, and channel intermediate molecules as diverse as fatty aldehydes and NADH co-factors. The design flexibility of RNA offered us an opportunity to study the effect of changing enzyme orientations on pathway yield. We presented a model that explains our experimental results wherein we see geometrically dependent metabolite channeling on *in vivo* RNA scaffolds.

We also showed that electronic flow in cyanobacteria can be rewired for biological $H_2$ production. We could demonstrate both light dependent hydrogen synthesis and hydrogen dependent cell growth. The latter can be accessed for directed evolution of new enzymatic properties and pathways. We end with showing that RNA scaffolds can also be expressed in cyanobacteria, enabling the combination of spatial metabolic control with renewable, photo-biological production.

Metabolic engineering is entering a new era propelled by cheaper gene-synthesis technologies[173], a better understanding of metabolic regulation[174], tools for spatial control[175], and an advanced ability to design enzyme catalysis[11]. Additionally, *in silico* metabolic models are more rigorous than ever[8]. Thus, biorefineries, which are already affecting substantial global impact[176] can be expected to play a more significant role across manufacturing industries in the future[177,178].

RNA is a powerful tool to synthetic biologists. RNA scaffolds can be composed of many structural, dynamic and functional regions. As we have shown, the structure of RNA can be designed predictably, and there are several assays being developed for proper assembly. In addition, recent advances in DNA construction[173,179] have made *in vivo* testing of designs faster and more reliable.

Synthetic biologists are constantly seeking to increase the complexity of their devices. RNA synthetic biology is offering tools to enable such control[180]. One such goal is the construction of synthetic ribosomes[181], capable of orthogonally incorporating non-natural amino acids wherein altered tRNA-protein interactions enable an expanded genetic code[182]. RNA scaffolds are also being employed to devise more precise genome editing tools[183]. Therapeutic applications of *in vivo* RNA scaffolds include functionalizing natural RNA scaffolds to enable drug delivery or isolation of pure samples. Similar developments in the fields of DNA packaging and origami for drug delivery[184,28] could offer strong synergistic opportunities for clinically applicable technologies to be implemented. More generally, the ability to simulate and predict the dynamics of structure-receptor binding interactions should enhance the design of such therapeutics[116].

Moving forward, innovations in high throughput design, synthesis and assaying functions for RNA structures will enable a greater range of applications to be developed. Developments in chip-based synthesis[179,185], new structure assembly assays such as SHAPE-seq[186], and improved genetically encodable electron microscopy tags[187,188] are enabling unprecedented scales of RNA synthesis, characterization and testing. Thus, the discovery of a variety of natural RNA structures and functions, an ever increasing understanding of how such features can be designed and an ability to rapidly implement and test ideas are indicators of a significant role for RNA scaffolds in future synthetic biology applications.

# A

# Supplementary information for *in vivo* RNA scaffold use

This appendix contains supplemental information* for experiments and designs described in Chapters 2 and 3. It presents detailed sequence information about the new

*The material in this section also forms the supporting online material for the published work

| RBD | Amino acid Sequence | Aptamer RNA sequence |
|---|---|---|
| RevN7D | TRQARRDRRRRWRERQR | GACAAGUUGGUCCGCACAGUUGCGAGGUGU |
| RevR11 | TRQARRNRRRQWRERQR | CGCUUAUGGUCAUUGAGUAUCUCCUGCCGA |
| BMV | KMTRAQRRAAARRNRWTAR | GGCGGUGGGUUUGGAAAACGGUAACGGGCA |
| Lambda | MDAQTRRRERRAEKQAQWK | GGGCCUGAAGAAGGGCCC |
| BIV-Tat | SGPRPRGTRGKGRRIRR | GGCUCGUGUAGCUCAUUAGCUCCGAGCC |
| HTLV-1 Rex | MPKTRRRPRRSQRKRP | GCUCAGGUCGAGGTACGCAAGTACCUCCCUUG GAGC |

**Figure A.1:** Sequence information for new aptamer-RNA binding domain pairs

aptamers tested and used on RNA scaffolds, sequence information about the RNA scaffold strands themselves, a list of plasmids that the were characterized, and information about the development of models shown in Figs. 3.7 and 3.8.

## RNA Binding domain-aptamer sets

The sequences for arginine-rich, short RNA binding peptide domains and their corresponding RNA ligands[97] added to our toolkit of aptamers for use on *in vivo* RNA scaffolds are shown in Fig A.1.

The new RNA binding domain-aptamer sets are mutually orthogonal, as seen from the table of $K_d$ values in A.2 of RNA binding domains (RBDs) tested *in vitro* and corresponding target aptamers (as reported in Bayer et al[97]).

We picked BIV-Tat from this list since it has an X-ray crystal structure reported[100] and showed high affinity to its aptamer in the above reports.

## Cloning and expression strategy

The following table lists the plasmid systems used for expression of the various components in *E. coli*.

| | RevN7D | RevR11Q | BMV Gag | Lambda N | BIV Tat | HTLV-1 Rex |
|---|---|---|---|---|---|---|
| **anti-RevN7D** | 120 | >1024 | >1024 | N.T. | >1024 | >1024 |
| **anti-RevR11Q** | >1024 | 100 | >1024 | N.T. | >1024 | >1024 |
| **Anti-BMV** | >1024 | >1024 | 20 | N.T. | >1024 | >1024 |
| **Lambda box** | >1024 | >1024 | >1024 | 90 | >1024 | >1024 |
| **BIV TAR** | >1024 | >1024 | >1024 | >1024 | 60 | >1024 |
| **HTLV XBE** | >1024 | >1024 | >1024 | >1024 | >1024 | 270 |

**Figure A.2:** Table showing $K_d$ values for new RNA binding domain-aptamer sets. First row contains the names of RNA binding domains (RBD). The first column is the list of the corresponding aptamers. Each entry in the table represents the tested dissociation constants ($K_d$) in nM, as tested and reported in Bayer et al [97]. N.T.: not tested.

The following tables list DNA sequences used for expression of various RNA scaffolds and

The Do, and D1 correspond to sequences that form discrete scaffolds, and one-dimensional polymerizing scaffolds respectively, as described in the main text and by Delebecque et al [13,85]. The two-dimensional polymerizing scaffolds are formed by two different strands, 2D_A and 2D_B, tiling together in the pattern shown in Fig 2.1. The geometric variants for the stem loop presenting the anti-BIV-Tat aptamer were created as shown with the 1-11 base-pairs added to the original 2D_B design stem loop.

## Model development for flux transfer on scaffolds

The scaffold for co-localization of AAR and DCB used in this work is based on the 2D design of Delebecque et al. [13], with the difference that the BIV-Tat aptamer now replaces the MS2 aptamer used in that work.† This change is not expected to affect the overall scaffold structure, as it lies outside the sequences that mediate scaffold

†The model developed in this section was worked out and explained with help from Jeffrey Way.

| Plasmid Name | Vector Backbone (resistance) | Expressed sequences |
|---|---|---|
| pPSGS088 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ RevN7D |
| pPSGS089 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ RevR11Q |
| pPSGS090 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ BMV Gag |
| pPSGS091 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ Lambda N |
| pPSGS092 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ BIV-Tat |
| pPSGS093 | pACYCDuet1 (Cm) | GFPA_($Gly_4SerGly_4$)_ HTLV-1 Rex |
| pPSGS094 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ RevN7D |
| pPSGS095 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ RevR11Q |
| pPSGS096 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ BMV Gag |
| pPSGS097 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ Lambda N |
| pPSGS098 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ BIV-Tat |
| pPSGS099 | pCOLADuet1 (Kan) | GFPB_($Gly_4SerGly_4$)_ HTLV-1 Rex |
| pPSGS100 | pACYCDuet1 (Cm) | BIVTAT_(6xHis)_SerGlySer_AAR (6xHis)_SerGlySer_PP7_($Gly_4SerGly_4$)_ADO |
| pPSGS139 (Fig S1) | pACYCDuet1 (Cm) | BIVTAT_(6xHis)_SerGlySer_AAR (Strep-Tag)_SerGlySer_PP7_($Gly_4SerGly_4$)_ADO |
| pPSGS083 pPSGS087, pPSGS118 pPSGS125-135 | pETDuet1 (Amp) | Please see Scaffold Expression table for details of relevant DNA sequences expressed (2D scaffold vector shown in Fig S2) |
| pPSGS153 | pETDuet1 (Amp) | 2D scaffold with anti-RevR11Q and MS2 |
| pPSGS144 | pACYCDuet1 (Cm) | PP7-PYC, mdh-BIV-TAT, lambda-FDH |
| pPSGS146 | pETDuet1 (Amp) | 2D scaffold with PP7, anti-BIV-TAT, and anti-lambdaN |
| pPSGS151 | pACYCDuet1 (Cm) | RevR11Q-eCA, PP7-PYC, mdh-BIV-TAT, lambda-FDH |
| pPSGS152 | pETDuet1 (Amp) | 2D scaffold with annti-RevR11Q, PP7, anti-BIV-TAT, and anti-lambdaN |

**Figure A.3:** List of plasmids used for Chapters 2 and 3.

| Scaffold Expression | DNA sequence |
|---|---|
| D0<br>(BIV-Tat and PP7)<br>pPSGS083 | GGAGGACTC**GGCTCGTGTAGCTCATTAGCTCCGAGCC**<br>GAGTCCTCGAATACGAGCTG**GGCACAGAAGATATGGCTTCGTGCC**<br>CAGGAAGTGTTCGCACTTCTCTCGTATTCGATTCCC_<T7TERM> |
| D1<br>(BIV-Tat and PP7)<br>pPSGS087 | GGGTAGGCGCCTAGCCTAATGTACATTAAGTTATTTTTCCGGATGAATAG<br>AATATATTCTAATAACGCAGGACTC<br>**GGCTCGTGTAGCTCATTAGCTCCGAGCC**<br>GAGTCCTCGAATACGAGCTG**GGCACAGAAGATATGGCTTCGTGCC**<br>CAGGAAGTGTTCGCACTTCTCTCGTATTCGATTGCG_<T7TERM> |
| 2D_A<br>(PP7)<br>pPSGS118 | GGGTCAGGAATCCTCCTGATAGCTATTTGGACAATTACGTACGTAGTTGA<br>TGACAACTACATGAAAATAAG**GGCACAGAAGATATGGCTTCGTGCC**<br>CTCTAGA_<T7TERM> |
| 2D_B<br>(BIV-Tat)<br>pPSGS118 | GGGTAGTTGTTATGGATTCCTGATTTATGGGACCCT<br>**GGCTCGTGTAGCTCATTAGCTCCGAGCC**<br>AGGGTCC_ ACTAGT_<T7TERM> |
| 2D_BIV_variants<br>(2D_B)<br>pPSGS125-135<br>(Default 7bp stem) | GGGTAGTTGTTATGGATTCCTGATTTATG***GGACCCT_<Rota-Variant>***<br>**_ GGCTCGTGTAGCTCATTAGCTCCGAGCC _*<Rota-Var-ReverseComplement>*_ *AGGGTCC*_** ACTAGT_<T7TERM> |
| Rota-Variant-8bp stem | G |
| Rota-Variant-9bp stem | TA |
| Rota-Variant-10bp stem | GGA |
| Rota-Variant-11bp stem | UGGC |
| Rota-Variant-12bp stem | AGGGA |
| Rota-Variant-13bp stem | GAGACC |
| Rota-Variant-14bp stem | AGUGAUG |
| Rota-Variant-15bp stem | GUUCCACG |
| Rota-Variant-16bp stem | AGUUCGUGA |
| Rota-Variant-17bp stem | GCAGUUCAAA |
| Rota-Variant-18bp stem | UGGUCAUCAAA |
| PP7 aptamer | GGCACAGAAGATATGGCTTCGTGCC |
| MS2 aptamer | CCACAGTCACTGGG |
| T7Term | CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG |

**Figure A.4:** DNA sequences used for expression of various RNA scaffolds

| RT-PCR primer | Primer sequence |
|---|---|
| 2D_F | GGGTCAGGAATCCTCCTGATAGC |
| 2D_R | GGTTATGCTAGTCTAGAGGGCACG |
| GAPA_F | ACTGACTGGTATGGGGTTCC |
| GAPA_R | AGGTTTAACGGCAGCTTTGA |

**Figure A.5:** Primers used to measure scaffold RNA levels

assembly. The RNA structure is expected to form a unit tile of the form $A_2B_2$. Six-base pair single-stranded sticky-end extensions emerge from the corners of the unit tile that allow connection to other tiles. The overall structure of assembled unit tiles is planar[13]. The RNA aptamers and T7 terminator element emerge from the 3' ends of the 6-bp extension and will be forced out of the plane of the unit tiles (Fig 2.1). Polymerization of the unit tiles via the 6-bp extensions is expected to force the PP7 and BIV-TAT aptamers that emerge from a given base-paired extension to be on the same side of the scaffold plane. This is because six base pairs correspond to about $\frac{1}{2}$ of a turn of an RNA double helix; the aptamers are attached to opposite strands. Thus, at the corners of the unit tiles, the two different aptamers will be in close proximity. (Figs 2.1, 3.7A) In principle, long-range interactions across the unit tile are also possible. For example, the assembly pattern in Fig 2.1 places diagonally related aptamers on one face of the scaffold plane and others on the opposite face. This configuration assumes a planar configuration around the Holliday junctions in the central RNA segments, and is based on: (1) an axis of 180 rotational symmetry around a y-axis drawn through the center of the structure, and (2) the shorter RNA, with the sequence UAGUUGUUAUGGAUUCCUGAUUUAUG-BIVTAT, has 26 bases in the segment within the scaffold, corresponding to about $2\frac{1}{2}$ turns of a helix; so that the aptamers at the beginning and the end of such a segment would be expected to be on opposite faces of the scaffold plane. Alternative configurations are possible; for example, the RNA Holliday junctions could be tetrahedral rather than planar, and the RNA segments passing through the Holliday junctions could have additional base-pairs per turn, changing the relative emergence sites of aptamers in different parts of the scaffold.

The representations shown in Fig 3.7 and 3.8 represent the above established arrangement using PDB files, 2QUX[113], 2A9X[114], and 4KVQ[115] from the NCBI protein

structure database. We have also used the Make-NA tool (James Stroud 2004, 2011) available on the University of Southern California servers for constructing the PDB representations of our stem and T7 terminator hairpins. The software Autodesk Maya (student version 2014) was used to render the 3D representation, along with a Molecular Maya toolkit (Digizyme).

# References


[1] David Fell and Athel Cornish-Bowden. *Understanding the control of metabolism*, volume 2. Portland press London, 1997.

[2] Priscilla E M Purnick and Ron Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, June 2009.

[3] W C Ruder, T Lu, and J J Collins. Synthetic Biology Moving into the Clinic. *Science*, 333(6047):1248–1252, September 2011.

[4] Yaakov Benenson. Biomolecular computing systems: principles, progress and potential. *Nature Reviews Genetics*, 13(7):455–468, June 2012.

[5] A J Ragauskas. The Path Forward for Biofuels and Biomaterials. *Science*, 311(5760):484–489, January 2006.

[6] S Pacala. Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies. *Science*, 305(5686):968–972, August 2004.

[7] Sung Kuk Lee, Howard Chou, Timothy S Ham, Taek Soon Lee, and Jay D Keasling. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, 19(6):556–563, December 2008.

[8] Jan Schellenberger, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, Aarash Bordbar, Nathan E Lewis, Sorena Rahmanian, Joseph Kang, Daniel R Hyduke, and Bernhard Ø Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–1307, August 2011.

[9] Jie Yuan, Bryson D Bennett, and Joshua D Rabinowitz. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature Protocols*, 3(8):1328–1340, July 2008.

[10] Christopher M Pirie, Marjan De Mey, Kristala L Jones Prather, and Parayil Kumaran Ajikumar. Integrating the protein and metabolic engineering toolkits for next-generation chemical biosynthesis. *ACS Chem. Biol.*, 8(4):662–672, April 2013.

[11] U T Bornscheuer, G W Huisman, R J Kazlauskas, S Lutz, J C Moore, and K Robins. Engineering the third wave of biocatalysis. *Nature*, 485(7397):185–194, May 2012.

[12] Hanson Lee, William C DeLoache, and John E Dueber. Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering*, 14(3):242–251, May 2012.

[13] C J Delebecque, A B Lindner, P. A. Silver, and F A Aldaye. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*, 333(6041):470–474, July 2011.

[14] John E Dueber, Gabriel C Wu, G Reza Malmirchegini, Tae Seok Moon, Christopher J Petzold, Adeeti V Ullal, Kristala L J Prather, and Jay D Keasling. Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology*, 27(8):753–759, August 2009.

[15] Farren J Isaacs, Daniel J Dwyer, and James J Collins. RNA synthetic biology. *Nature Biotechnology*, 24(5):545–554, May 2006.

[16] S J Culler, K G Hoff, and C D Smolke. Reprogramming Cellular Behavior with RNA Controllers Responsive to Endogenous Proteins. *Science*, 330(6008):1251–1255, November 2010.

[17] Lei S Qi, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, 152(5):1173–1183, February 2013.

[18] Annette Khaled, Songchuan Guo, Feng Li, and Peixuan Guo. Controllable Self-Assembly of Nanoparticles for Specific Delivery of Multiple Therapeutic Molecules to Cancer Cells Using RNA Nanotechnology. *Nano Letters*, 5(9):1797–1808, September 2005.

[19] Faisal A Aldaye, William T Senapedis, Pamela A Silver, and Jeffrey C Way. A Structurally Tunable DNA-Based Extracellular Matrix. *Journal of the American Chemical Society*, 132(42):14727–14729, October 2010.

[20] Christina M Agapakis, Patrick M Boyle, and Pamela A Silver. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology*, 8(6):527–535, May 2012.

[21] Todd O. Yeates, Cheryl A Kerfeld, Sabine Heinhorst, Gordon C Cannon, and Jessup M Shively. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nature reviews. Microbiology*, 6(9):681–691, September 2008.

[22] P. A. Srere. Complexes of sequential metabolic enzymes. *Annual Review of Biochemistry*, 56(1):89–124, 1987.

[23] J T Penrod and J R Roth. Conserving a Volatile Metabolite: a Role for Carboxysome-Like Organelles in Salmonella enterica. *Journal of bacteriology*, 188(8):2865–2874, April 2006.

[24] Stuart Smith and Shiou-Chuan Tsai. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports*, 24(5):1041–1072, 2007.

[25] Todd O. Yeates, Christopher S. Crowley, and Shiho Tanaka. Bacterial microcompartment organelles: protein shell structure and evolution. *Annual review of biophysics*, 39:185–205, June 2010.

[26] Anna H Chen and Pamela A Silver. Designing biological compartmentalization. *Trends in Cell Biology*, 22(12):662–670, December 2012.

[27] Michael Erkelenz, Chi-Hsien Kuo, and Christof M Niemeyer. DNA-Mediated Assembly of Cytochrome P450 BM3 Subdomains. *Journal of the American Chemical Society*, 133(40):16111–16118, October 2011.

[28] Jinglin Fu, Minghui Liu, Yan Liu, Neal W Woodbury, and Hao Yan. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *Journal of the American Chemical Society*, 134(12):5516–5519, March 2012.

[29] Mingxu You, Ruo-Wen Wang, Xiaobing Zhang, Yan Chen, Kelong Wang, Lu Peng, and Weihong Tan. Photon-Regulated DNA-Enzymatic Nanostructures by Molecular Assembly. *ACS Nano*, 5(12):10090–10095, December 2011.

[30] Tae Seok Moon, John E. Dueber, Eric Shiue, Kristala L. Jones Prather, and Kristala L. Prather. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering*, 12(3):298–305, May 2010.

[31] R J Conrado, G C Wu, J T Boock, H Xu, S Y Chen, T Lebar, J Turnsek, N Tomsic, M Avbelj, R Gaber, T Koprivnjak, J Mori, V Glavnik, I Vovk, M Bencina, V Hodnik, G Anderluh, J E Dueber, R Jerala, and M P DeLisa. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic acids research*, 40(4):1879–1889, February 2012.

[32] Ido Golding and Edward C Cox. RNA dynamics in live Escherichia coli cells. *PNAS*, 101(31):11310–11315, August 2004.

[33] L Felipe Barros and Cristián Martínez. An Enquiry into Metabolite Domains. *Biophysical Journal*, 92(11):3878–3884, June 2007.

[34] Ronny Lorenz, Stephan HF Bernhart, Christian Hoener Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB*, 6(1):26, 2011.

[35] Nicholas R Markham and Michael Zuker. UNAFold: software for nucleic acid folding and hybridization. *Methods in molecular biology (Clifton, N.J.)*, 453:3–31, 2008.

[36] Jessica S Reuter and David H Mathews. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11(1):129, 2010.

[37] Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, November 2010.

[38] Neocles B Leontis, Aurelie Lescoute, and Eric Westhof. The building blocks and motifs of RNA architecture. *Current Opinion in Structural Biology*, 16(3):279–287, June 2006.

[39] Luc Jaeger and Arkadiusz Chworos. The architectonics of programmable RNA and DNA nanostructures. *Current opinion in structural biology*, 16(4):531–543, August 2006.

[40] José Almeida Cruz and Eric Westhof. The Dynamic Landscapes of RNA Architecture. *Cell*, 136(4):604–609, February 2009.

[41] Ignacio Tinoco, Jr and Carlos Bustamante. How RNA folds. *Journal of Molecular Biology*, 293(2):271–281, October 1999.

[42] Lauren S Waters and Gisela Storz. Regulatory RNAs in Bacteria. *Cell*, 136(4):615–628, February 2009.

[43] Wade C Winkler and Ronald R Breaker. Regulation of bacterial gene expression by riboswitches. *Annual review of microbiology*, 59:487–517, 2005.

[44] Evgeny Nudler and Alexander S Mironov. The riboswitch control of bacterial metabolism. *Trends in biochemical sciences*, 29(1):11–17, January 2004.

[45] F F Lim, T P TP Downey, and D S DS Peabody. Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. *Journal of Biological Chemistry*, 276(25):22507–22513, June 2001.

[46] I Hirao, M Spingola, D Peabody, and A D Ellington. The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Molecular diversity*, 4(2):75–89, 1998.

[47] Gary W Witherell and Olke C Uhlenbeck. Specific RNA binding by Q.beta. coat protein. *Biochemistry (Washington)*, 28(1):71–76, January 1989.

[48] P Guo, S Erickson, and D Anderson. A small viral RNA is required for in vitro packaging of bacteriophage phi 29 DNA. *Science*, 236(4802):690–694, May 1987.

[49] A A Simpson, Y Tao, P G Leiman, M O Badasso, Y He, P J Jardine, N H Olson, M C Morais, S Grimes, D L Anderson, T S Baker, and M G Rossmann. Structure of the bacteriophage phi29 DNA packaging motor. *Nature*, 408(6813):745–750, December 2000.

[50] Chao-Zhou Ni, Rashid Syed, Ramadurgam Kodandapani, John Wickersham, David S Peabody, and Kathryn R Ely. Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure*, 3(3):255–263, March 1995.

[51] David S Peabody and Kathryn R Ely. Control of translational repression by protein - protein interactions. *Nucleic acids research*, 20(7):1649–1655, 1992.

[52] Peixuan Guo, Chunlin Zhang, Chaoping Chen, Kyle Garver, and Mark Trottier. Inter-RNA Interaction of Phage $\phi$29 pRNA to Form a Hexameric Complex for Viral DNA Transportation. *Molecular cell*, 2(1):149–155, July 1998.

[53] Jason G Underwood, Andrew V Uzilov, Sol Katzman, Courtney S Onodera, Jacob E Mainzer, David H Mathews, Todd M Lowe, Sofie R Salama, and David Haussler. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nature Methods*, 7(12):995–1001, November 2010.

[54] Michael Kertesz, Yue Wan, Elad Mazor, John L Rinn, Robert C Nutter, Howard Y Chang, and Eran Segal. Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 467(7311):103–107, September 2010.

[55] Tim R Mercer and John S Mattick. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*, 20(3):300–307, March 2013.

[56] M C Tsai, O Manor, Y Wan, N Mosammaparast, J K Wang, F Lan, Y Shi, E Segal, and H Y Chang. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, 329(5992):689–693, August 2010.

[57] David C Zappulla and Thomas R Cech. Yeast telomerase RNA: a flexible scaffold for protein subunits. *PNAS*, 101(27):10024–10029, July 2004.

[58] Carla A Theimer and Juli Feigon. Structure and function of telomerase RNA. *Current opinion in structural biology*, 16(3):307–318, June 2006.

[59] Farren J Isaacs, Daniel J Dwyer, Chunming Ding, Dmitri D Pervouchine, Charles R Cantor, and James J Collins. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7):841–847, June 2004.

[60] Maung Nyan Win and Christina D Smolke. Targeted cleavage: Tuneable cis-cleaving ribozymes. 104(38):14881–14882, September 2007.

[61] Travis S Bayer and Christina D Smolke. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology*, 23(3):337–343, February 2005.

[62] Chase L Beisel, Travis S Bayer, Kevin G Hoff, and Christina D Smolke. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Molecular Systems Biology*, 4, October 2008.

[63] Kirill A Afonin, Eckart Bindewald, Alan J Yaghoubian, Neil Voss, Erica Jacovetty, Bruce A Shapiro, and Luc Jaeger. In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nature Nanotechnology*, 5(9):676–682, August 2010.

[64] Arkadiusz Chworos, Isil Severcan, Alexey Y Koyfman, Patrick Weinkam, Emin Oroudjev, Helen G Hansma, and Luc Jaeger. Building programmable jigsaw puzzles with RNA. *Science*, 306(5704):2068–2072, December 2004.

[65] Sergey M Dibrov, Jaime McLean, Jerod Parsons, and Thomas Hermann. Self-assembling RNA square. *Proceedings of the National Academy of Sciences*, 108(16):6405–6408, 2011.

[66] Isil Severcan, Cody Geary, Arkadiusz Chworos, Neil Voss, Erica Jacovetty, and Luc Jaeger. A polyhedron made of tRNAs. *Nature Chemistry*, 2(9):772–779, July 2010.

[67] Edouard Bertrand, Pascal Chartrand, Matthias Schaefer, Shailesh M Shenoy, Robert H Singer, and Roy M Long. Localization of ASH1 mRNA Particles in Living Yeast. *Molecular cell*, 2(4):437–445, October 1998.

[68] A S Brodsky and P. A. Silver. Pre-mRNA processing factors are required for nuclear export. *Rna*, 6(12):1737–1749, December 2000.

[69] Dahlene Fusco, Nathalie Accornero, Brigitte Lavoie, Shailesh M Shenoy, Jean-Marie Blanchard, Robert H Singer, and Edouard Bertrand. Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells. *Current Biology*, 13(2):161–167, January 2003.

[70] Martina Schifferer and Oliver Griesbeck. Application of aptamers and autofluorescent proteins for RNA visualization. *Integrative Biology*, 1(8-9):499–505, 2009.

[71] Thuc T Le, Sébastien Harlepp, Călin C Guet, Kimberly Dittmar, Thierry Emonet, Tao Pan, and Philippe Cluzel. Real-time RNA profiling within a single bacterium. In *Proceedings of the National Academy of Sciences*, pages 9160–9164. Department of Biochemistry and Molecular Biology, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, June 2005.

[72] Kenneth C Keiler. RNA localization in bacteria. *Current Opinion in Microbiology*, 14(2):155–159, April 2011.

[73] Natalia E Broude. Analysis of RNA localization and metabolism in single live bacterial cells: achievements and challenges. *Molecular Microbiology*, 80(5):1137–1147, April 2011.

[74] Maria Valencia-Burton, Ron M McCullough, Charles R Cantor, and Natalia E Broude. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature Methods*, 282(5387):296–298, April 2007.

[75] Takeaki Ozawa, Yutaka Natori, Moritoshi Sato, and Yoshio Umezawa. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nature Methods*, April 2007.

[76] A D Ellington and J W Szostak. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287):818–822, August 1990.

[77] Craig Tuerk and Larry Gold. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968):505–510, 1990.

[78] David S Wilson and Jack W Szostak. In vitro selection of functional nucleic acids. *Annual Review of Biochemistry*, 68(1):611–647, November 1999.

[79] Hung-Wei Yiu, Vadim V Demidov, Paul Toran, Charles R Cantor, and Natalia E Broude. RNA Detection in live bacterial cells using fluorescent protein complementation triggered by interaction of two RNA aptamers with two RNA-binding peptides. *Pharmaceuticals*, 4(3):494–508, 2011.

[80] M Valencia-Burton, A Shah, J Sutin, A Borogovac, R M McCullough, C R Cantor, A Meller, and N E Broude. Spatiotemporal patterns and transcription kinetics of induced RNA in single bacterial cells. *Proceedings of the National Academy of Sciences*, 106(38):16399–16404, September 2009.

[81] Songchuan Guo, Nuska Tschammer, Sulma Mohammed, and Peixuan Guo. Specific delivery of therapeutic RNAs to cancer cells via the dimerization mechanism of phi29 motor pRNA. *Human gene therapy*, 16(9):1097–1109, September 2005.

[82] Luc Ponchon and Frédéric Dardel. Recombinant RNA technology: the tRNA scaffold. *Nature Methods*, 4(7):571–576, June 2007.

[83] L Ponchon, M Catala, B Seijo, M El Khouri, F Dardel, S Nonin-Lecomte, and C Tisne. Co-expression of RNA-protein complexes in Escherichia coli and applications to RNA biology. *Nucleic acids research*, June 2013.

[84] Minghui Liu, Jinglin Fu, Christian Hejesen, Yuhe Yang, Neal W Woodbury, Kurt Gothelf, Yan Liu, and Hao Yan. A DNA tweezer-actuated enzyme nanoreactor. *Nature Communications*, 4:2127, July 2013.

[85] Camille J Delebecque, Pamela A Silver, and Ariel B Lindner. Designing and using RNA scaffolds to assemble proteins in vivo. *Nature Protocols*, 7(10):1797–1807, September 2012.

[86] Daniel C Ducat, Gairik Sachdeva, and Pamela A Silver. Rewiring hydrogenase-dependent redox circuits in cyanobacteria. *Proceedings of the National Academy of Sciences*, 108(10):3941–3946, March 2011.

[87] A Schirmer, M A Rude, X Li, E Popova, and S B del Cardayre. Microbial Biosynthesis of Alkanes. *Science*, 329(5991):559–562, July 2010.

[88] Ailen M Sánchez, George N Bennett, and Ka-Yiu San. Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. *Metabolic Engineering*, 7(3):229–239, May 2005.

[89] Luisa S Gronenberg, Ryan J Marcheschi, and James C Liao. Next generation biofuel engineering in prokaryotes. *Current Opinion in Chemical Biology*, pages 1–10, April 2013.

[90] Eugenia M Clerico, Jayna L Ditty, and Susan S Golden. Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. In *Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria*, pages 155–171. Humana Press, Totowa, NJ, 2007.

[91] Daniel C. Ducat, Jeffrey C. Way, and Pamela A. Silver. Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology*, 29(2):95–103, February 2011.

[92] James B McKinlay and Caroline S Harwood. Photobiological production of hydrogen gas as a biofuel. *Current Opinion in Biotechnology*, 21(3):244–251, June 2010.

[93] Rosalind E Franklin and Raymond G Gosling. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, 172:156–157, 1953.

[94] James D Watson and Francis HC Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[95] Faisal A Aldaye and Hanadi F Sleiman. Supramolecular DNA nanotechnology. *Pure and Applied Chemistry*, 81(12):2157–2181, 2009.

[96] Gairik Sachdeva, Abhishek Garg, David Godding, Jeffrey C Way, and Pamela A Silver. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. *Nucleic acids research*, 42(14):9493–9503, October 2014.

[97] Travis S Bayer, Lauren N Booth, Scott M Knudsen, and Andrew D Ellington. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *Rna*, 11(12):1848–1857, December 2005.

[98] M Spingola and D S Peabody. MS2 coat protein mutants which bind Qbeta RNA. *Nucleic acids research*, 25(14):2808–2815, July 1997.

[99] Maria Valencia-Burton and Natalia E Broude. Visualization of RNA using fluorescence complementation triggered by aptamer-protein interactions (RFAP) in live bacterial cells. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]*, Chapter 17:Unit 17.11, December 2007.

[100] J D Puglisi, L Chen, S Blanchard, and A D Frankel. Solution Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex. *Science*, 270(5239):1200–1203, November 1995.

[101] Oleksii Sliusarenko, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Molecular Microbiology*, 80(3):612–627, March 2011.

[102] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, June 2012.

[103] Ralf Jungmann, Maier S Avendaño, Johannes B Woehrstein, Mingjie Dai, William M Shih, and Peng Yin. Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nature Methods*, 11(3):313–318, February 2014.

[104] J S Paige, K Y Wu, and S R Jaffrey. RNA Mimics of Green Fluorescent Protein. *Science*, 333(6042):642–646, July 2011.

[105] Richard P Feynman. There's plenty of room at the bottom. *Engineering and Science*, 23(5):22–36, 1960.

[106] Ofer Idan and Henry Hess. The Origins of Activity Enhancement in Enzyme Cascades on Scaffolds. *ACS Nano*, 7(10):1309051410010 09, October 2013.

[107] Douglas M Warui, Ning Li, Hanne Nørgaard, Carsten Krebs, J Martin Bollinger, and Squire J Booker. Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase. *Journal of the American Chemical Society*, 133(10):3316–3319, March 2011.

[108] G N Vemuri, M A Eiteman, and E Altman. Effects of Growth Mode and Pyruvate Carboxylase on Succinic Acid Production by Metabolically Engineered Strains of Escherichia coli. *Applied and environmental microbiology*, 68(4):1715–1727, April 2002.

[109] Susana J Berríos-Rivera, George N Bennett, and Ka-Yiu San. Metabolic engineering of Escherichia coli: increase of NADH availability by overexpressing an NAD(+)-dependent formate dehydrogenase. *Metabolic Engineering*, 4(3):217–229, July 2002.

[110] S Reiser and C Somerville. Isolation of mutants of Acinetobacter calcoaceticus deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *Journal of bacteriology*, 179(9):2969–2975, May 1997.

[111] Alan T Bull. The renaissance of continuous culture in the post-genomics age. *Journal of industrial microbiology & biotechnology*, 37(10):993–1021, October 2010.

[112] B F Eichman, J M Vargason, B H M Mooers, and P S Ho. The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions. *Proceedings of the National Academy of Sciences*, 97(8):3971–3976, April 2000.

[113] Jeffrey A Chao, Yury Patskovsky, Steven C Almo, and Robert H Singer. Structural basis for the coevolution of a viral RNA-protein complex. *Nature Structural & Molecular Biology*, 15(1):103–105, January 2008.

[114] Thomas C Leeper, Zafiria Athanassiou, Ricardo L A Dias, John A Robinson, and Gabriele Varani. TAR RNA Recognition by a Cyclic Peptidomimetic of Tat Protein †,‡. *Biochemistry (Washington)*, 44(37):12362–12372, September 2005.

[115] Basile Khara, Navya Menon, Colin Levy, David Mansell, Debasis Das, E Neil G Marsh, David Leys, and Nigel S Scrutton. Production of Propane and Other Short-Chain Alkanes by Structure-Based Engineering of Ligand Specificity in Aldehyde-Deformylating Oxygenase. *ChemBioChem*, 14(10):1204–1208, June 2013.

[116] Avi Robinson-Mosher, Tamar Shinar, Pamela A Silver, and Jeffrey Way. Dynamics simulations for engineering macromolecular interactions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(2):025110, 2013.

[117] Dan Wang, Qiang Li, Wangliang Li, Jianmin Xing, and Zhiguo Su. Improvement of succinate production by overexpression of a cyanobacterial carbonic anhydrase in Escherichia coli. *Enzyme and microbial technology*, 45(6-7):491–497, December 2009.

[118] Ralph E H Sims, Warren Mabee, Jack N Saddler, and Michael Taylor. Bioresource Technology. *Bioresource Technology*, 101(6):1570–1580, March 2010.

[119] G Charles Dismukes, Damian Carrieri, Nicholas Bennette, Gennady M Ananyev, and Matthew C Posewitz. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Current Opinion in Biotechnology*, 19(3):235–240, June 2008.

[120] Stephanie G Hays and Daniel C Ducat. Engineering cyanobacteria as photosynthetic feedstock factories. *Photosynthesis Research*, pages 1–11, February 2014.

[121] Maria L Ghirardi, Matthew C Posewitz, Pin-Ching Maness, Alexandra Dubini, Jianping Yu, and Michael Seibert. Hydrogenases and Hydrogen Photoproduction in Oxygenic Photosynthetic Organisms *. *Annual Review of Plant Biology*, 58(1):71–91, June 2007.

[122] Maria Lucia Ghirardi, Alexandra Dubini, Jianping Yu, and Pin-Ching Maness. Photobiological hydrogen-producing systems. *Chemical Society reviews*, 38(1):52–61, 2008.

[123] Oliver Schmitz, Gudrun Boison, Ralf Hilscher, Barbara Hundeshagen, Wolfgang Zimmer, Fnedrich Lottspeich, and Hermann Bothe. Molecular Biological Analysis of a Bidirectional Hydrogenase from Cyanobacteria. *European Journal of Biochemistry*, 233(1):266–276, October 1995.

[124] S T Stripp, G Goldet, C Brandmayr, O Sanganas, K A Vincent, M Haumann, F A Armstrong, and T Happe. How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proceedings of the National Academy of Sciences*, 106(41):17331–17336, October 2009.

[125] T K Antal, T E Krendeleva, T V Laurinavichene, V V Makarova, M L Ghirardi, A B Rubin, A A Tsygankov, and M Seibert. The dependence of algal H2 production on Photosystem II and O2 consumption activities in sulfur-deprived Chlamydomonas reinhardtii cells. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1607(2-3):153–160, December 2003.

[126] O Kruse, J Rupprecht, K P Bader, S Thomas-Hall, P M Schenk, G Finazzi, and B Hankamer. Improved Photobiological H2 Production in Engineered Green Algal Cells. *Journal of Biological Chemistry*, 280(40):34170–34177, September 2005.

[127] A Hansel and P Lindblad. Towards optimization of cyanobacteria as biotechnologically relevant producers of molecular hydrogen, a clean and renewable energy source. *Applied microbiology and biotechnology*, 50(2):153–160, August 1998.

[128] P W King, M C Posewitz, M L Ghirardi, and M Seibert. Functional Studies of [FeFe] Hydrogenase Maturation in an Escherichia coli Biosynthetic System. *Journal of bacteriology*, 188(6):2163–2172, March 2006.

[129] Günter A Peschek. Electron transport reactions in respiratory particles of hydrogenase-induced Anacystis nidulans. *Archives of Microbiology*, 125(1-2):123–131, March 1980.

[130] Christina M Agapakis, Daniel C Ducat, Patrick M Boyle, Edwin H Wintermute, Jeffrey C Way, and Pamela A Silver. Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of Biological Engineering*, 4(1):3, 2010.

[131] Matthew C Posewitz, Sharon L Smolinski, Saradadevi Kanakagiri, Anastasios Melis, Michael Seibert, and Maria L Ghirardi. Hydrogen photoproduction is attenuated by disruption of an isoamylase gene in Chlamydomonas reinhardtii. *The Plant Cell Online*, 16(8):2151–2163, 2004.

[132] H Niederholtmeyer, B T Wolfstadter, D F Savage, P. A. Silver, and J C Way. Engineering Cyanobacteria To Synthesize and Export Hydrophilic Products. *Applied and environmental microbiology*, 76(11):3462–3466, May 2010.

[133] Robert J Porra. The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. In *link.springer.com.ezp-prod1.hul.harvard.edu*, pages 633–640. Springer-Verlag, Berlin/Heidelberg, 2005.

[134] Anja Hemschemeier, Swanny Fouchard, Laurent Cournac, Gilles Peltier, and Thomas Happe. Hydrogen production by Chlamydomonas reinhardtii: an elaborate interplay of electron sources and sinks. *Planta*, 227(2):397–407, September 2007.

[135] Jens Rupprecht, Ben Hankamer, Jan H Mussgnug, Gennady Ananyev, Charles Dismukes, and Olaf Kruse. Perspectives and advances of biological H2 production in microorganisms. *Applied microbiology and biotechnology*, 72(3):442–449, August 2006.

[136] T K Antal and P Lindblad. Production of H2 by sulphur-deprived cells of the unicellular cyanobacteria Gloeocapsa alpicola and Synechocystis sp. PCC 6803 during dark incubation with methane or at various extracellular pH. *Journal of Applied Microbiology*, 98(1):114–120, January 2005.

[137] Vincent Chochois, David Dauvillée, Audrey Beyly, Dimitri Tolleter, Stéphan Cuiné, Hélène Timpano, Steven Ball, Laurent Cournac, and Gilles Peltier. Hydrogen production in Chlamydomonas: photosystem II-dependent and-independent pathways differ in their requirement for starch metabolism. *Plant physiology*, 151(2):631–640, 2009.

[138] O Schmitz and H Bothe. The diaphorase subunit HoxU of the bidirectional hydrogenase as electron transferring protein in cyanobacterial respiration? *Naturwissenschaften*, 83(11):525–527, November 1996.

[139] Günter A Peschek. Anaerobic hydrogenase activity in Anacystis nidulans H2-dependent photoreduction and related reactions. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 548(2):187–202, November 1979.

[140] Toshiharu Shikanai. Cyclic Electron Transport Around Photosystem I: Genetic Approaches. *Annual Review of Plant Biology*, 58(1):199–217, June 2007.

[141] Olivier Guerrini, Bénédicte Burlat, Christophe Léger, Bruno Guigliarelli, Philippe Soucaille, and Laurence Girbal. Characterization of Two 2[4Fe4S] Ferredoxins from Clostridium acetobutylicum. *Current Microbiology*, 56(3):261–267, December 2007.

[142] Yasuo Asada, Yoji Koike, Jörg Schnackenberg, Masato Miyake, Ieaki Uemura, and Jun Miyake. Heterologous expression of clostridial hydrogenase in the cyanobacterium Synechococcus PCC7942. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1490(3):269–278, February 2000.

[143] M Winkler, S Kuhlgert, M Hippler, and T Happe. Characterization of the Key Step for Light-driven Hydrogen Evolution in Green Algae. *Journal of Biological Chemistry*, 284(52):36620–36627, December 2009.

[144] K McNeely, Y Xu, N Bennette, D A Bryant, and G C Dismukes. Redirecting Reductant Flux into Hydrogen Production via Metabolic Engineering of Fermentative Carbon Metabolism in a Cyanobacterium. *Applied and environmental microbiology*, 76(15):5032–5038, July 2010.

[145] R Surzycki, L Cournac, G Peltier, and J D Rochaix. Potential for hydrogen production with inducible chloroplast gene expression in Chlamydomonas. *Proceedings of the National Academy of Sciences*, 104(44):17548–17553, October 2007.

[146] Anastasios Melis, John Neidhardt, and John R Benemann. Dunaliella salina (Chlorophyta) with small chlorophyll antenna sizes exhibit higher photosynthetic productivities and photon use efficiencies than normally pigmented cells - Springer. *Journal of Applied Phycology*, 10(6):515–525, 1998.

[147] Tatyana V Laurinavichene, Alexander S Fedorov, Maria L Ghirardi, Michael Seibert, and Anatoly A Tsygankov. Demonstration of sustained hydrogen photoproduction by immobilized, sulfur-deprived Chlamydomonas reinhardtii cells. *International Journal of Hydrogen Energy*, 31(5):659–667, 2006.

[148] Sergey Kosourov, Michael Seibert, and Maria L Ghirardi. Effects of extracellular pH on the metabolic pathways in sulfur-deprived, H2-producing Chlamydomonas reinhardtii cultures. *Plant & cell physiology*, 44(2):146–155, February 2003.

[149] I Phillips and P Silver. A new biobrick assembly strategy designed for facile protein engineering. *2006*, January 2006.

[150] Carol R Andersson, Nicholas F Tsinoremas, Jeffrey Shelton, Nadya V Lebedeva, Justin Yarrow, Hongtao Min, and Susan S Golden. Application of bioluminescence to the study of circadian rhythms in cyanobacteria. *Methods in enzymology*, 305:527–542, 2000.

[151] R John Ellis. Biochemistry: Tackling unintelligent design. *Nature*, 463(7278):164–165, January 2010.

[152] Rowan F Sage, Tammy L Sage, and Ferit Kocacinar. Photorespiration and the Evolution of C 4Photosynthesis. *Annual Review of Plant Biology*, 63(1):19–47, June 2012.

[153] Murray R Badger and G Dean Price. CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *Journal of Experimental Botany*, 54(383):609–622, February 2003.

[154] G Dean Price, John R Coleman, and Murray R Badger. Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium Synechococcus PCC7942. *Plant physiology*, 100(2):784–793, 1992.

[155] Jeffrey C Cameron, Steven C Wilson, Susan L Bernstein, and Cheryl A Kerfeld. Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell*, 155(5):1131–1140, November 2013.

[156] Anna H Chen, Avi Robinson-Mosher, David F Savage, Pamela A Silver, and Jessica K Polka. The Bacterial Carbon-Fixing Organelle Is Formed by Shell Envelopment of Preassembled Cargo. *PLoS ONE*, 8(9):e76127, September 2013.

[157] Walter Bonacci, Poh K Teng, Bruno Afonso, Henrike Niederholtmeyer, Patricia Grob, Pamela A Silver, and David F Savage. Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences*, 109(2):478–483, January 2012.

[158] Niall M Mangan, Michael P Brenner, and Ron Milo. Systems analysis of the CO2 concentrating mechanism in cyanobacteria. *eLife*, 3:e02043, April 2014.

[159] Rui Sousa, Debabrata Patra, and Eileen M Lafer. Model for the mechanism of bacteriophage T7 RNAP transcription initiation and termination. *Journal of Molecular Biology*, 224(2):319–334, March 1992.

[160] G D Price, S M Howitt, K Harrison, and M R Badger. Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain PCC7942

involved in carboxysome assembly and function. *Journal of bacteriology*, 175(10):2871–2879, 1993.

[161] Thomas C Taylor, Anders Backlund, Karin Bjorhall, Robert J Spreitzer, and Inger Andersson. First crystal structure of Rubisco from a green alga, Chlamydomonas reinhardtii. *Journal of Biological Chemistry*, 276(51):48159–48164, 2001.

[162] B B Menon, S Heinhorst, J M Shively, and G C Cannon. The Carboxysome Shell Is Permeable to Protons. *Journal of bacteriology*, 192(22):5881–5886, October 2010.

[163] E E Soltes-Rak, M E ME Mulligan, and J R JR Coleman. Identification and characterization of a gene encoding a vertebrate-type carbonic anhydrase in cyanobacteria. *J Bacteriol*, 179(3):769–774, February 1997.

[164] Gordon C. Cannon, Sabine Heinhorst, and Cheryl A. Kerfeld. Carboxysomal carbonic anhydrases: Structure and role in microbial CO2 fixation. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1804(2):382–392, February 2010.

[165] B. M. Long, M. R. Badger, S. M. Whitney, and G. D. Price. Analysis of Carboxysomes from Synechococcus PCC7942 Reveals Multiple Rubisco Complexes with Carboxysomal Proteins CcmM and CcaA. *Journal of Biological Chemistry*, 282(40):29323–29335, August 2007.

[166] Jessica K Polka and Pamela A Silver. Building synthetic cellular organization. *Molecular biology of the cell*, 24(23):3585–3587, 2013.

[167] C A Kerfeld and D Loque. Design and Implementation of Novel and/or Enhanced Bacterial Microcompartments for Customizing Metabolism. US Patent Office, 2012.

[168] G Dean Price, Jasper JL Pengelly, Britta Forster, Jiahui Du, Spencer M Whitney, Susanne von Caemmerer, Murray R Badger, Susan M Howitt, and John R Evans. The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. *Journal of Experimental Botany*, 64(3):753–768, 2013.

[169] Charles Darwin. On the origin of species by means of natural selection. *Murray. London*, 1871.

[170] D G Gibson, J I Glass, C Lartigue, V N Noskov, R Y Chuang, M A Algire, G A Benders, M G Montague, L Ma, M M Moodie, C Merryman, S Vashee, R Krishnakumar, N Assad-Garcia, C Andrews-Pfannkoch, E A Denisova, L Young,

Z Q Qi, T H Segall-Shapiro, C H Calvey, P P Parmar, C A Hutchison, H O Smith, and J C Venter. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 329(5987):52–56, July 2010.

[171] C Adami, C Ofria, and T C Collier. Evolution of biological complexity. *Proceedings of the National Academy of Sciences*, 97(9):4463–4468, April 2000.

[172] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, November 2004.

[173] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, April 2009.

[174] Luca Gerosa and Uwe Sauer. Regulation and control of metabolic fluxes in microbes. *Current Opinion in Biotechnology*, 22(4):566–575, August 2011.

[175] Ofer Idan and Henry Hess. Engineering enzymatic cascades on nanoscale scaffolds. *Current Opinion in Biotechnology*, 24(4):606–611, August 2013.

[176] Chris J Paddon and Jay D Keasling. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nature Publishing Group*, 12(5):355–367, April 2014.

[177] Vikramaditya G Yadav, Marjan De Mey, Chin Giaw Lim, Parayil Kumaran Ajikumar, and Gregory Stephanopoulos. The future of metabolic engineering and synthetic biology: Towards a systematic practice. *Metabolic Engineering*, 14(3):233–241, May 2012.

[178] Vikramaditya G Yadav and Gregory Stephanopoulos. Metabolic Engineering: The Ultimate Paradigm for Continuous Pharmaceutical Manufacturing. *ChemSusChem*, pages n/a–n/a, April 2014.

[179] Sriram Kosuri, Nikolai Eroshenko, Emily M LeProust, Michael Super, Jeffrey Way, Jin Billy Li, and George M Church. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature biotechnology*, 28(12):1295–1299, November 2010.

[180] Joe C Liang, Ryan J Bloom, and Christina D Smolke. Engineering Biological Systems with Synthetic RNA Molecules. *Molecular Cell*, 43(6):915–926, September 2011.

[181] Kaihang Wang, Heinz Neumann, Sew Y Peak-Chew, and Jason W Chin. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature Biotechnology*, 25(7):770–777, June 2007.

[182] Heinz Neumann, Kaihang Wang, Lloyd Davis, Maria Garcia-Alai, and Jason W Chin. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, 464(7287):441–444, February 2010.

[183] P Mali, L Yang, K M Esvelt, J Aach, M Guell, J E DiCarlo, J E Norville, and G M Church. RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121):823–826, February 2013.

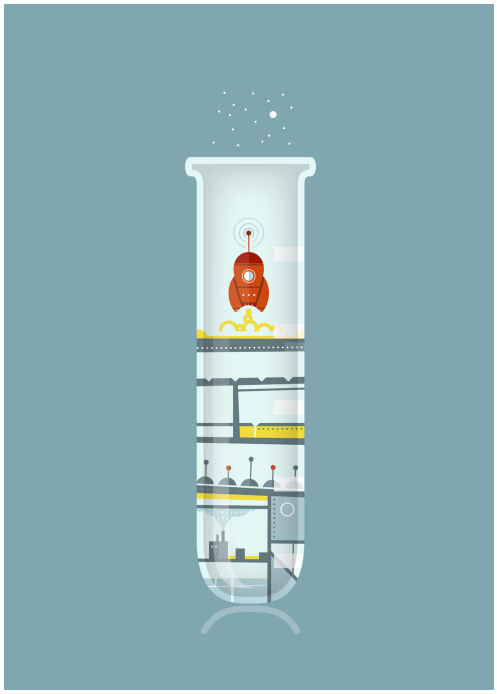
[184] S M Douglas, I Bachelet, and G M Church. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science*, 335(6070):831–834, February 2012.

[185] Cheng-Hsien Wu, Matthew R Lockett, and Lloyd M Smith. RNA-Mediated Gene Assembly from DNA Arrays. *Angewandte Chemie International Edition*, 51(19):4628–4632, March 2012.

[186] Julius B Lucks, Stefanie A Mortimer, Cole Trapnell, Shujun Luo, Sharon Aviran, Gary P Schroth, Lior Pachter, Jennifer A Doudna, and Adam P Arkin. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences*, 108(27):11063–11068, May 2011.

[187] Xiaokun Shu, Varda Lev-Ram, Thomas J Deerinck, Yingchuan Qi, Ericka B Ramko, Michael W Davidson, Yishi Jin, Mark H Ellisman, and Roger Y Tsien. A Genetically Encoded Tag for Correlated Light and Electron Microscopy of Intact Cells, Tissues, and Organisms. *PLoS Biology*, 9(4):e1001041, April 2011.

[188] Jeffrey D Martell, Thomas J Deerinck, Yasemin Sancak, Thomas L Poulos, Vamsi K Mootha, Gina E Sosinsky, Mark H Ellisman, and Alice Y Ting. engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature biotechnology*, pages 1–9, October 2012.



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (XII) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.